THE UNIVERSITY
OF MICHIGAN

SEP 12 1962

ENGINEERING
LIBRARY

АВТОМАТИКА
И
ТЕЛЕМЕХАНИКА

Vol. 22, No. 8, August, 1961

Translation Published February, 1962

SOVIET INSTRUMENTATION AND
CONTROL TRANSLATION SERIES

Automation and Remote Control

(The Soviet Journal *Avtomatika i Telemekhanika* in English Translation)

■ This translation of a Soviet journal on automatic control is published as a service to American science and industry. It is sponsored by the Instrument Society of America under a grant in aid from the National Science Foundation, continuing a program initiated by the Massachusetts Institute of Technology.

SOVIET INSTRUMENTATION AND CONTROL TRANSLATION SERIES

Instrument Society of America

Philip A. Sprague
President
Dr. Ralph H. Tripp
Past President
Nathan Cohn
President-elect-Secretary
Henry J. Noebels
Dept. Vice President
Dr. Benjamin W. Thomas
Dept. Vice President-elect
E. A. Adler
Dept. Vice President
Francis S. Hoag
Dept. Vice President
John J. McDonald
Dept. Vice President
John R. Mahoney
Dept. Vice President-elect
John C. Koch
Treasurer
Alonzo B. Parsons
Dist. I Vice President
H. Kirk Fallin
Dist. II Vice President
Harold J. Bowman
Dist. III Vice President
Fred R. Gilmer
Dist. IV Vice President
Charles H. Callier
Dist. V Vice President
Otto J. Lessa
Dist. VI Vice President
J. Howard Park, III
Dist. VII Vice President
C. Roy Horton
Dist. VIII Vice President
Max Curtis
Dist. IX Vice President
Kenneth S. Vriesen
Dist. X Vice President
Allan E. Lee
Dist. XI Vice President

International Headquarters

William H. Kushnick
Executive Director
Charles W. Covey
Editor, ISA Journal
Herbert S. Kindler
Director, Tech. & Educ. Services

ISA Publications Committee

Charles O. Badgett, *Chairman*

Jere E. Brophy	George A. Larsen	Joshua Stern
Dr. Enoch J. Durbin	Thomas G. MacAnespie	Frank S. Swaney
Prof. Richard W. Jones	John E. Read	Richard A. Terry

Translations Advisory Board of the Publications Committee

Jere E. Brophy, *Chairman*

T. J. Higgins	S. G. Eskin	G. Werbizky

■ This translation of the Soviet Journal *Avtomatika i Telemekhanika* is published and distributed at nominal subscription rates under a grant in aid to the Instrument Society of America from the National Science Foundation. This translated journal, and others in the Series (see back cover), will enable American scientists and engineers to be informed of work in the fields of instrumentation, measurement techniques, and automatic control reported in the Soviet Union.

The original Russian articles are translated by competent technical personnel. The translations are on a cover-to-cover basis and the Instrument Society of America and its translators propose to translate faithfully all of the scientific material in *Avtomatika i Telemekhanika*, permitting readers to appraise for themselves the scope, status, and importance of the Soviet work. All views expressed in the translated material are intended to be those of the original authors and not those of the translators nor the Instrument Society of America.

Publication of *Avtomatika i Telemekhanika* in English translation started under the present auspices in April, 1958, with Russian Vol. 18, No. 1 of January, 1957. The program has been continued with the translation and printing of the 1958-1961 issues.

Transliteration of the names of Russian authors follows the system known as the British Standard. This system has recently achieved wide adoption in the United Kingdom, and is currently being adopted by a large number of scientific journals in the United States.

Readers are invited to submit to the Instrument Society of America comments on the quality of the translations and the content of the articles. Pertinent correspondence will be published in the Society's monthly publication, the ISA JOURNAL. Space will also be made available in the ISA JOURNAL for such replies as may be received from Russian authors to comments or questions by the readers.

1961 Volume 22 Subscription Prices:

Per year (12 issues), starting with Vol. 22, No. 1

General:	*United States and Canada*	$35.00
	Elsewhere	38.00
Libraries of nonprofit academic institutions:		
	United States and Canada	$17.50
	Elsewhere	20.50
Single issues to everyone, each		$ 6.00

1957 Volume 18, 1958 Volume 19, 1959 Volume 20, and 1960 Volume 21 issues also available. Prices upon request.

See back cover for combined subscription to entire Series.

Subscriptions and requests for information on back issues should be addressed to the:

Instrument Society of America
530 William Penn Place, Pittsburgh 19, Penna.

Translated and printed by Consultants Bureau Enterprises, Inc.

Copyright © 1962 by the Instrument Society of America

Automation and Remote Control

A translation of Avtomatika i Telemekhanika, a publication of the Academy of Sciences of the USSR

EDITORIAL BOARD OF AVTOMATIKA I TELEMEKHANIKA

D. I. Ageikin
M. A. Aizerman
A. B. Chelyustkin *(Assoc. Editor)*
E. G. Dudnikov
N. Ya. Festa
V. A. Ilin
A. G. Iosuf'yan
V. V. Karibskii
A. V. Khramoi
B. Ya. Kogan
V. S. Kulebakin
S. A. Lebedev
A. Ya. Lerner
A. M. Letov *(Assoc. Editor)*
V. S. Malov
B. N. Petrov
Yu. P. Portnov-Sokolov
B. S. Sotskov
A. A. Tal' *(Corresp. Secretary)*
V. A. Trapeznikov *(Editor in Chief)*
Ya. Z. Tsypkin
G. M. Ulanov
A. A. Voronov
S. V. Yablonskii

Vol. 22, No. 8 Russian Original Dated August, 1961 February, 1962

CONTENTS

	PAGE	RUSS. PAGE
Absolute Stability of Nonlinear Systems of Automatic Control. V. M. Popov	857	961
The Condition Governing the Absence of Over-Regulation in Certain Nonlinear Automatic Control Systems. G. M. Ostrovskii	876	980
L. S. Pontryagin's Maximum Principle and Optimal Progamming of Rocket Thrust. V. K. Isaev	881	986
Automatic Optimization of Space Distribution. III. L. N. Fitsner	894	1001
The Determination of Optimal Characteristics for an Extremal System with Random Disturbances. N. V. Grishko	904	1013
On the Qustion of Choice of Parameters in the Control Part of a Gradient-Type Automatic Optimization System. T. I. Tovstykha	918	1027
A System for Automatic Programming for Machine Tools. B. G. Tamm	929	1038
Automatic Thickness Control of Cold-Rolled Strip, Based on Edge-Thickness Measurements. Ch'ü Shou-tê	943	1055
Some Applications of Contact Grids. Antonin Svoboda	947	1061
On the Theory of Switching Network Synthesis. F. Svoboda	956	1071
On Digital Reproduction of Signals in Analog Telemetering Systems. Ya. A. Kupershmidt	965	1080
The Investigation of Branched Power Nets as Communication and Remote Control Channels. A. A. Il'in	972	1088
Automatic Remote Control of Water Distribution. I. B. Chizh	978	1095
Analytic and Structural Description of Mechanical Transmissions in Automatic Control Systems with Restrictions and Backlash. Ts. G. Litovchenko and Yu. P. Yakovenko	982	1100
Estimating the Reliability of Automatic Systems from the Results of Testing an Incomplete Set of Equipment. S. M. Kuznetsov	990	1108
A Fast Multichannel Distributor Using Transistors. Ya. E. Belen'ky and V. N. Mikhailovsky	999	1117
The Stability of Single-Stage Aperiodic Low-Frequency Vacuum Tube Amplifiers. V. I. Pampuro	1004	1123
Pneumatic Integrator with Aperiodic Unit Throttle in Low Pressure Area. Ch'i Ju-hsien	1013	1133
A. A. Andronov (On the 60th Anniversary of His Birthday)	1018	1139

SOVIET INSTRUMENTATION AND CONTROL TRANSLATION SERIES

Automation and Remote Control

A translation of Avtomatika i Telemekhanika, a publication of the Academy of Sciences of the USSR

EDITORIAL BOARD OF AVTOMATIKA I TELEMEKHANIKA

B. I. Antonov	A. Ya. Lerner	A. A. Tal'
M. A. Aizerman	A. M. Lerov	(German Democratic Rep.)
A. A. Chelyustkin	(Assoc. Editor)	V. A. Trapeznikov
(Assoc. Editor)	V. V. Solodov	(Editor-in-Chief)
E. G. Shelikhov	(Assoc. Editor)	Ya. Z. Tsypkin
M. Ye. Perel'	V. A. Kotel'nikov	G. M. Ulanov
	S. Ya. Khaikin	A. A. Voronov
	V. S. Kulebakin	S. V. Yablonskii
	Ya. I. Khurgin	
	S. S. Lavrov	
	B. S. Sotskov	

Vol. 22, No. 8 — Initial Publication Dated August, 1961 — February, 1962

CONTENTS

PAGE / PAGE

Analytic Stability of Nonlinear Systems of Automatic Control. V. M. Popov ..... 1013 / 1019

The Operating Characteristics of Two-Position Regulator in Certain Nonlinear Automatic Control Systems. B. N. Petrov ..... 1019 / 1026

The Optimum Problem of Optimal Programming and Optimal Programming of Rocket Thrust ..... 1026 / 1034

Automatic Optimization of Static Characteristics. M. A. Rozenblat ..... 1034 / 1041

The Optimum Problem of Optimal Characteristics for an Internal System with Random Disturbances. B. N. Petrov ..... 1041 / 1048

On the Question of Choice of Parameters of the Control Plan of a Gradient-Type Automatic Optimizing System. T. A. Tikhomirova ..... 1048 / 1053

A Device for Automatic Programming of Angular Deviations. S. G. Tarmin ..... 1053 / 1059

Automatic Control of Certain Objects Having Several Variables ..... 1059 / 1067

On the Theory of Selecting Automatic Systems. F. Gerchels ..... 1067 / 1073

On the Optimal Approximation of Signals in Automatic Tracking Systems. Ya. A. Kupershtein ..... 1073 / 1080

Simultaneous Determination of Maximal Power Gain and Noise Control Channels. A. A. Ilin ..... 1080 / 1083

Automatic Remote Control of Water Distribution. I. A. Chikin ..... 1083 / 1088

Analysis and Structural Description of Discharge Transitions in Automatic Control Systems with Servitization and Relevant. Ya. G. Lipovchenko and Yu. P. Yakovenko ..... 1088 / 1100

Eliminating the Reliability of Automatic Systems from the Results of Testing an Incomplete Set of Equipment. S. M. Kuznetsov ..... 1100 / 1108

A Fast Multichannel Discharge Transition. Ya. E. Belen'kii and V. N. Mikhailovskii ..... 1108 / 1117

The Stability of Single-Stage Aperiodic Low-Frequency Vacuum Tube Amplifiers. V. I. Pavlov ..... 1117 / 1123

Automatic Passage with Aperiodic Intermediate Drive Through Areas. CAN'T READ ..... 1123 / 1129

A. A. Andronov (On the 60th Anniversary of His Birthday) ..... 1129 / 1139

ABSOLUTE STABILITY OF NONLINEAR SYSTEMS OF AUTOMATIC CONTROL

V. M. Popov (Bucharest)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 961-979, August, 1961
Original article submitted January 17, 1961

The problem of absolute stability of an "indirect control" system with a single nonlinearity is investigated by using a method which differs from the second method of Lyapunov. The main condition of the obtained criterion of absolute stability is expressed in terms of the transfer function of the system linear part. It is also shown that by forming the standard Lyapunov function -"a quadratic form plus the integral of the nonlinearity" it is not possible in the case considered here to obtain a wider stability domain than the one obtained from the presented criterion. Graphical criteria of absolute continuity are also given by means of the phase-amplitude characteristic or by what is known as the modified phase-amplitude characteristic" of the system linear part.

In the present paper the absolute stability is investigated of nonlinear systems of "indirect control." The existing literature in this field (see [1] for example) deals exclusively with a direct application of Lyapunov's method. In this paper the solution is obtained by a different method, and this enables one to get new results.

It is assumed that the reader is not familiar with the author's previous publications. Therefore, not only the most general results are given here but also a very simple example shall be considered.

By using a new method the author has also investigated the absolute stability of other types of systems of differential equations (for example, of the system of "direct control"), as well as of other classes of nonlinear functions (for example, of functions whose graph is contained within a sector). In all these cases the absolute stability of the system with several nonlinearities is also studied (the case of systems with many controlling devices).

In his most recent papers now in press, the author has studied the stability in certain critical cases and also the stability of systems of differential equations with an "aftereffect".

1. Statement of the Problem

Systems of "indirect control" are considered which can be described by the following system of differential equations:

$$\frac{dx_l}{dt} = \sum_{k=1}^{n} a_{lk}x_k + b_l\varphi(\sigma) \qquad (l = 1, 2, \ldots, n), \tag{1.1}$$

$$\frac{d\xi}{dt} = \varphi(\sigma), \tag{1.2}$$

$$\sigma = \sum_{l=1}^{n} c_l x_l - \gamma\xi, \tag{1.3}$$

where a_{lk}, b_l, c_l, and γ are constant, and $\varphi(\sigma)$ is a function of class A*, that is, a continuous function satisfying the condition

$$\varphi(0) = 0 \tag{1.4}$$

and also the inequality

* It is assumed that the quantity σ_* introduced in [6] always vanishes. It should just be mentioned that if $\sigma_* \neq 0$, the trivial solution of the system (1.1)-(1.3) cannot be asymptotically stable.

$$\varphi(\sigma)\,\sigma > 0 \quad \text{for } \sigma \neq 0 \tag{1.5}$$

The system (1.1)-(1.3) admits the trivial solution

$$x_1 = \xi = \sigma = 0, \tag{1.6}$$

whose stability is under investigation.

It is assumed that the trivial solution of the linear system with constant coefficients

$$\frac{dy_l}{dt} = \sum_{k=1}^{n} a_{lk} y_k \qquad (l = 1, 2, \ldots, n), \tag{1.7}$$

where a_{lk} are the same as in (1.1), is asymptotically stable, or (which is equivalent) that all real parts of the eigenvalues of the matrix (a_{lk}) are negative. The conditions are being sought which would be satisfied by the quantities a_{lk}, b_l, c_l, and γ in order that the trivial solution of the system (1.1)-(1.3) be asymptotically stable, whatever the function $\varphi(\sigma)$ of the class A (in other words, a condition of asymptotic absolute stability of the trivial solution). As we know to achieve this, it is necessary that•

$$\gamma > 0. \tag{1.8}$$

We shall therefore consider in the sequel the inequality (1.8) to be satisfied.

2. Introductory Definitions. Formulation of Criterion of Absolute Stability

Consider the functions $\psi_{lm}(t)$ ($l = 1, 2, \ldots, n$; $m = 1, 2, \ldots, n$), defined when $t \geq 0$, and being the solutions of the equations

$$\frac{d\psi_{lm}(t)}{dt} = \sum_{k=1}^{n} a_{lk}\psi_{km}(t) \qquad \begin{pmatrix} l = 1, 2, \ldots, n \\ m = 1, 2, \ldots, n \end{pmatrix} \tag{2.1}$$

together with the initial conditions

$$\psi_{lm}(0) = \delta_{lm} \quad (l = 1, 2, \ldots, n;\ m = 1, 2, \ldots, n), \tag{2.2}$$

where $\delta_{lm} = 0$ when $l \neq m$, and $\delta_{lm} = 1$ when $l = m$.

The functions $\psi_{lm}(t)$ form the fundamental system of solutions for the system (1.7).

Let $x_l(t)$, $\xi(t)$ be the solution of the equations (1.1)-(1.3)•• which satisfies the initial conditions $x_l(0) = x_{l0}$, $\xi(0) = \xi$, and let $\varphi[\sigma(t)]$ be a function of t obtained by substituting the function $\sigma(t) = \sum c_l x_l(t) - \gamma\xi(t)$ in $\varphi(\sigma)$ [see (1.3)].

As a solution of the system (1.1) we obtain

$$x_l(t) = \sum_{m=1}^{n} \psi_{lm}(t)\, x_{m0} + \sum_{m=1}^{n} \int_0^t \psi_{lm}(t-\zeta)\, b_m \varphi(\sigma(\zeta))\, d\zeta. \tag{2.3}$$

It follows from (1.3) that

$$\sigma(t) = \sum_{m=1}^{n} \mu_m(t)\, x_{m0} - \int_0^t \nu(t-\zeta)\,\varphi(\sigma(\zeta))\, d\zeta - \gamma\xi(t), \tag{2.4}$$

•If $\gamma < 0$, the trivial solution is unstable when $\varphi(\sigma) = h\sigma$, $h > 0$. If $\gamma = 0$, the condition $\xi(t) = 0$ is not satisfied for all solutions of the system.

•• The existence of solutions is a consequence of the assumptions made in section 1. But their uniqueness is not assumed in the sequel. A solution can always be extended in view of the conditions for stability as formulated below.

where

$$\mu_m(t) = \sum_{l=1}^{n} c_l \psi_{lm}(t), \tag{2.5}$$

$$\nu(t) = -\sum_{l=1}^{n}\sum_{m=1}^{n} c_l \psi_{lm}(t)\, b_m. \tag{2.6}$$

In agreement with our assumptions the trivial solution of the system (1.7) is asymptotically stable, and therefore two positive constants K_0 and K_1 can be found such that [see (2.1) and (2.2)] for all $t \geq 0$ the inequality

$$|\psi_{lm}(t)| < K_1 e^{-K_0 t}, \qquad t \geqslant 0 \qquad \begin{pmatrix} l = 1, 2, \ldots, n, \\ m = 1, 2, \ldots, n \end{pmatrix}. \tag{2.7}$$

takes place.

Therefore, for $\nu(t)$ [see (2.6)] the inequality

$$|\nu(t)| < K_2 e^{-K_0 t} \quad (t \geqslant 0), \tag{2.8}$$

is valid where

$$K_2 = \sum_{l=1}^{n} |c_l| \sum_{l=1}^{n} |b_l| K_1$$

We note that in view of the inequality (2.7), the Fourier transform of the function $\nu(t)$ given by the formula

$$N(j\omega) = \int_0^\infty e^{-j\omega t} \nu(t)\, dt \qquad (j = \sqrt{-1}). \tag{2.9}$$

must exist.

We introduce the function

$$G(j\omega) = N(j\omega) + \frac{\gamma}{j\omega}, \tag{2.10}$$

which shall be called the transfer function of the system linear part (see Appendix 1).

Now the following theorem can be stated.

Theorem. If a nonnegative quantity q exists such that for all real w the inequality*

$$\mathrm{Re}\,(1 + j\omega q)\, G(j\omega) \geqslant 0, \tag{2.11}$$

takes place, then the trivial solution of the system (1.1)-(1.3) is asymptotically absolutely stable provided the assumptions made in Section 1 remain valid.

* ReX (or ImX) denotes the real (or, respectively, imaginary) part of a complex quantity X.

3. Proof of the Theorem in Section 2

Consider again the solution x_l (t), ξ (t), σ (t) of the system (1.1)-(1.3) as well as the function φ (σ (l)), corresponding to it. For any positive quantity T we define auxiliary functions

$$\varphi_T(t) = \begin{cases} \varphi(\sigma(t)) & \text{for } 0 \leqslant t \leqslant T, \\ 0 & \text{for } t > T, \end{cases} \tag{3.1}$$

$$\lambda_T(t) = -\int_0^t v(t-\zeta)\,\varphi_T(\zeta)\,d\zeta - q\int_0^t \frac{dv(t-\zeta)}{dt}\varphi_T(\zeta)d\zeta - q[v(0)+\gamma]\,\varphi_T(t), \tag{3.2}$$

where q is the quantity occurring in the theorem, and v (t) is given by the formula (2.6); it follows that [see (2.2)]

$$v(0) = -\sum_{l=1}^{n} c_l b_l. \tag{3.3}$$

It can easily be seen that the function λ_T (t) is bounded for all $0 \leq t \geq T$. When $t > T$, the inequality (see Appendix 1)

$$|\lambda_T(t)| < K_3 e^{-K_4 t} \quad (K_3 > 0,\ t > T), \tag{3.4}$$

takes place in which K_3 is independent of t. This guarantees the existence of the Fourier transform

$$L_T(j\omega) = \int_0^\infty e^{-j\omega t}\lambda_T(t)\,dt. \tag{3.5}$$

There exists also the transform

$$F_T(j\omega) = \int_0^\infty e^{-j\omega t}\varphi_T(t)\,dt. \tag{3.6}$$

In view of (2.6), (2.1), and (2.7) the Fourier transform of the function dv (t)/ dt exists and can be written as [see (2.9)]

$$\int_0^\infty e^{-j\omega t}\frac{dv(t)}{dt}dt = j\omega N(j\omega) - v(0); \tag{3.7}$$

Therefore, by taking the Fourier transform of (3.2) [see (3.5), (3.6), and (3.7)] we obtain

$$L_T(j\omega) = -[(1 + j\omega q)\,N(j\omega) + q\gamma]\,F_T(j\omega). \tag{3.8}$$

The following function of T is introduced:

$$\rho(T) = \int_0^\infty \lambda_T(t)\,\varphi_T(t)\,dt. \tag{3.9}$$

The Parseval formula can be applied in (3.9) because the function φ_T (t) is continuous for $0 \leq t \leq T$ and it vanishes for $t > T$, and the function λ_T (t) is also continuous for $0 \leq t \leq T$, and for $t > T$ the inequality (3.4) takes place. We now obtain

$$\rho(T) = \frac{1}{2\pi} \int_{-\infty}^{\infty} L_T(j\omega)\, \overline{F_T(j\omega)}\, d\omega, \tag{3.10}$$

where $\overline{F_T(j\omega)}$ is the conjugate complex of $F_T(j\omega)$.

By replacing $L_T(j\omega)$ by (3.8) in (3.10), and by taking into account that

$$F_T(j\omega)\, \overline{F_T(j\omega)} = |F_T(j\omega)|^2, \tag{3.11}$$

we obtain

$$\rho(T) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} |F_T(j\omega)|^2 [(1 + j\omega q) N(j\omega) + q\gamma]\, d\omega. \tag{3.12}$$

By comparing their real parts we obtain

$$\rho(T) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} |F_T(j\omega)|^2 \operatorname{Re}[(1 + j\omega q) N(j\omega) + q\gamma]\, d\omega. \tag{3.13}$$

Observing that [see (2.10) and (2.11)]

$$\operatorname{Re}[(1 + j\omega q) N(j\omega) + q\gamma] = \operatorname{Re}(1 + j\omega q)\left(N(j\omega) + \frac{\gamma}{j\omega}\right) = \operatorname{Re}(1 + j\omega q)\, g(j\omega) > 0, \tag{3.14}$$

gives the inequality

$$\rho(T) \leq 0, \tag{3.15}$$

By substituting (3.1) into (3.9) we obtain

$$\rho(T) = \int_0^T \lambda_T(t)\, \varphi(\sigma(t))\, dt < 0. \tag{3.16}$$

By substituting the expression for φ_T (t) into (3.2) and by applying the formula (2.4) as well as the formula

$$\frac{d\sigma(t)}{dt} = \sum_{m=1}^{n} \frac{d\mu_m(t)}{dt}\, x_{m0} - \int_0^t \frac{d\nu(t-\zeta)}{dt}\, \varphi(\sigma(\zeta))\, d\zeta - [\nu(0) + \gamma]\, \varphi(\sigma(t)), \tag{3.17}$$

easily following from (2.4), we obtain within the interval $0 \leq t \leq T$

$$\lambda_T(t) = \sigma(t) + q\frac{d\sigma(t)}{dt} + \gamma\xi(t) - \sum_{m=1}^{n} \left(\mu_m(t) + q\frac{d\mu_m(t)}{dt}\right) x_{m0}. \tag{3.18}$$

By substituting this expression into (3.16) we obtain

$$\int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt + q\int_0^T \varphi(\sigma(t))\,\frac{d\sigma(t)}{dt}\,dt + \gamma\int_0^T \varphi(\sigma(t))\,\xi(t)\,dt$$

$$-\sum_{m=1}^{n}\left[x_{m0}\int_0^T\left(\mu_m(t) + q\,\frac{d\mu_m(t)}{dt}\right)\varphi(\sigma(t))\,dt \leqslant 0. \right. \tag{3.19}$$

Each term of the expression (3.19) shall be considered in turn. For the first term we have the inequality [see (1.4) and (1.5)]

$$-\int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt \leqslant 0. \tag{3.20}$$

The second term can be rewritten with the aid of the identity

$$\int_0^T \varphi(\sigma(t))\,\frac{d\sigma(t)}{dt}\,dt = F(\sigma(T)) - F(\sigma(0)), \tag{3.21}$$

where

$$F(\sigma) = \int_0^{\sigma} \varphi(\sigma)\,d\sigma.$$

In the case when $\varphi(\sigma)$ belongs to the class A [see (1.4) and (1.5)], the condition

$$F(\sigma) \geq 0 \tag{3.22}$$

is fulfilled, with the equality only taking place when $\sigma = 0$. As q is nonnegative, one has, of course, the inequality

$$-qF(\sigma(T)) \leq 0. \tag{3.23}$$

The third term can be rewritten as follows [see (1.2)]:

$$\int_0^T \varphi(\sigma(t))\,\xi(t)\,dt = \int_0^T \frac{d\xi(t)}{dt}\,\xi(t)\,dt = \frac{1}{2}\,\xi^2(T) - \frac{1}{2}\,\xi_0^2. \tag{3.24}$$

As $\gamma > 0$ [see (1.8)] we have

$$-\frac{1}{2}\,\gamma\xi^2(T) \leqslant 0. \tag{3.25}$$

As far as the remaining terms are concerned we note [see Appendix 3] that there exists a positive quantity K_4 such that the inequality:

$$\left|\int_0^T\left(\mu_m(t) + q\,\frac{d\mu_m(t)}{dt}\right)\varphi(\sigma(t))\,dt\right| \leqslant \frac{1}{n}\,K_4\,\sup_{0<\zeta<T}|\xi(\zeta)| \quad (m = 1, 2, \ldots n). \tag{3.26}$$

takes place.

Using the identities (3.21) and (3.24) as well as the inequalities (3.26), one is able to rewrite the inequality (3.19) as

$$\int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt + qF(\sigma(T)) + \frac{1}{2}\gamma\xi^2(T)$$

$$-K_4 \max_{m=1,2,\ldots,n}(|x_{m0}|)\sup_{0<\zeta<T}|\xi(\zeta)| \leqslant qF(\sigma(0)) + \frac{1}{2}\gamma\xi_0^2. \tag{3.27}$$

We shall show that the trivial solution of the system (1.1)-(1.3) is stable. To this end we combine the inequalities (3.27), (3.20), and (3.23)

$$\frac{1}{2}\gamma\xi^2(T) - K_4 \max_{m=1,2,\ldots,n}(|x_{m0}|)\sup_{0<\zeta<T}|\xi(\zeta)| \leqslant qF(\sigma(0)) + \frac{1}{2}\gamma\xi_0^2. \tag{3.28}$$

Let t be an arbitrary positive quantity. The inequality (3.28) is valid for any $T > 0$. In particular, (3.25) remains valid for T (t) such that the inequality

$$0 \leq T(t) \leq t \tag{3.29}$$

takes place, and also the equality

$$|\xi(T(t))| = \sup_{0<\zeta<t}|\xi(\zeta)|. \tag{3.30}$$

Such a quantity T (t), of course, exists. It is easy to see that in view of (3.29) the inequalities

$$|\xi(T(t))| \leqslant \sup_{0<\zeta<T(t)}|\xi(\zeta)| \leqslant \sup_{0<\zeta<t}|\xi(\zeta)|. \tag{3.31}$$

take place.

By comparing (3.31) and (3.30) we obtain

$$|\xi(T(t))| = \sup_{0<\zeta<T(t)}|\xi(\zeta)|. \tag{3.32}$$

When considering (3.28) for T = T (t) and by using (3.32) we obtain

$$\frac{1}{2}\gamma\xi^2(T(t)) - K_4 \max_{m=1,2,\ldots,n}(|x_{m0}|)|\xi(T(t))| - qF(\sigma(0)) - \frac{1}{2}\gamma\xi_0^2 \leqslant 0. \tag{3.33}$$

The polynomial of the variable $|\xi(T(t))|$ in the lefthand side of the inequality (3.33) has real roots of opposite sign, when made equal to zero. The inequality (3.33) is equivalent to the inequality

$$|\xi(T(t))| \leqslant \frac{1}{\gamma}K_4 \max_{m=1,2,\ldots,n}(|x_{m0}|) + \frac{1}{\gamma}\sqrt{K_4^2\left(\max_{m=1,2,\ldots,n}(|x_{m0}|)\right)^2 + 2\gamma\left(qF(\sigma(0)) + \frac{1}{2}\gamma\xi_0^2\right)}, \tag{3.34}$$

which is valid for all $t > 0$.

One can find two positive quantities K_5 and K_6 [see Appendix 4] such that for all $t > 0$ the inequality

$$|x_i(t)| \leqslant K_5 \max_{m=1,2,\ldots,n}(|x_{m0}|) + K_6 \sup_{0<\zeta<t}|\xi(\zeta)|. \tag{3.35}$$

takes place.

By applying (3.30), (3.34), and (3.35) we finally obtain

$$|\xi(t)| \leqslant \frac{1}{\gamma} K_4 \max_{m=1,2,\ldots,n} (|x_{m0}|) + \frac{1}{\gamma} + \sqrt{K_4^2 (\max_{m=1,2,\ldots,n} (x_{m0}|))^2 + 2\gamma (qF(\sigma(0)) + \frac{1}{2}\gamma\xi_0^2)}, \tag{3.36}$$

$$|x_l(t)| \leqslant \left(K_5 + \frac{1}{\gamma} K_4 K_6\right) \max_{m=1,2,\ldots,n} (|x_{m0}|) + \frac{1}{\gamma} K_6 \sqrt{K_4^2 (\max_{m=1,2,\ldots,n} (|x_{m0}|))^2 + 2\gamma \left(qF(\sigma(0)) + \frac{1}{2}\gamma\xi_0^2\right)}. \tag{3.37}$$

Let ϵ be an arbitrary positive quantity. There exists than a $\delta > 0$ such that when $|x_{l0}| < \delta$ and $|\xi_0| < \delta$, the righthand sides of the inequalities (3.36) and (3.37) are less than ϵ;* and therefore the inequalities $|x_l(t)| < \epsilon$ and $|\xi(t)| < \epsilon$ take place. Consequently, the trivial solution of the system (1.1)-(1.3) is stable according to Lyapunov. Moreover, it follows from (3.36) and (3.37) that all the solutions of the system are bounded. This implies that all the solutions of the system (1.1)-(1.3) can be extended for all $t > 0$.

It shall be shown now that the trivial solution is also asymptotically stable for any $\varphi(\sigma)$ of the class A. We combine for this purpose the inequalities (3.27), (3.23), and (3.25):

$$\int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt \leqslant K_4 \max_{m=1,2,\ldots,n} (|x_{m0}|) \sup_{0\leqslant\zeta\leqslant T} |\xi(\zeta)| + qF(\sigma(0)) + \frac{1}{2}\gamma\xi_0^2. \tag{3.38}$$

By making use of (3.36)

$$\int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt \leqslant f_1(x_{m0}, \xi_0), \tag{3.39}$$

where $f_1(x_{m0}, \xi_0)$ is a specified function of the variables x_{m0} and ξ_0.

Taking into account the reaction [see (1.1), (1.3)]

$$\frac{d\sigma(t)}{dt} = \sum_{l=1}^n c_l \frac{dx_l}{dt} - \gamma \frac{d\xi}{dt} = \sum_{l=1}^n c_l \left(\sum_{k=1}^n a_{lk}x_k + b_l\varphi(\sigma)\right) - \gamma\varphi(\sigma), \tag{3.40}$$

[in the above formula x_k and ξ satisfy the inequalities (3.37) and (3.36), σ is determined by the equation (1.3) and $\varphi(\sigma)$ is a continuous function of σ], we obtain the inequality

$$\left|\frac{d\sigma(t)}{dt}\right| < f_2(x_{m0}, \xi_0), \tag{3.41}$$

where $f_2(x_{m0}, \xi_0)$ does not depend on t.

By using (3.39) and (3.41) we obtain [see Appendix 5]

$$\lim_{t\to\infty} \sigma(t) = 0 \tag{3.42}$$

* Here we make use of the fact that $F(\sigma(0))$ is a continuous function of $\sigma(0)$, vanishing when $\sigma(0) = 0$, and also of the equality $\sigma(0) = \sum_{l=1}^n c_l x_{l0} - \gamma\xi_0$ [see (1.3)].

and also

$$\lim_{t\to\infty} \varphi(\sigma(t)) = 0. \tag{3.43}$$

because $\varphi(\sigma)$ is a continuous function.

From (3.43) and from the equations (2.3) one obtains [see Appendix 6]

$$\lim_{t\to\infty} x_l(t) = 0. \tag{3.44}$$

The relation (1.3) implies [see (3.42) and (3.44)]

$$\lim_{t\to\infty} \xi(t) = \lim_{t\to\infty} \frac{1}{\tau}\left(\sum_{l=1}^{n} x_l(t) - \sigma t\right) = 0. \tag{3.45}$$

The theorem has thus been proved.

4. Comparing the Results with the Results Obtainable with the Aid of Lyapunov Functions of Known Kind

It is of interest to compare the criteria (2.11) of absolute stability with the ones which can be obtained by constructing a Lyapunov function of the kind described as a "quadratic form plus the integral of the nonlinearity" [6-8]. We shall show that the latter are included in the criteria of the theorem in section 2, that is, if for the system under investigation there exists a Lyapunov function of the above kind, then a nonnegative quantity q also exists such that the inequality (2.11) takes place.

By differentiating the equation (1.3) we obtain a system equivalent to the (1.1)-(1.3):

$$\begin{aligned} \frac{dx_l}{dt} &= \sum_{k=1}^{n} a_{lk}x_k + b_l\varphi(\sigma) \qquad (l = 1, 2, \ldots, n), \\ \frac{d\sigma}{dt} &= \sum_{l=1}^{n} c_l\left(\sum_{k=1}^{n} a_{lk}x_k + b_l\varphi(\sigma)\right) - \gamma\varphi(\sigma). \end{aligned} \tag{4.1}$$

It is shown in Appendix 7 that the most general form of a negative definite Lyapunov function which is of the kind described as a "quadratic form plus the integral of the nonlinearity" and which can be formed for the system (4.1) takes the form

$$V = \sum_{l=1}^{n}\sum_{m=1}^{n} x_l r_{lm} x_m - \alpha\left(\sigma - \sum_{l=1}^{n} c_l x_l\right)^2 - 2\beta\int_0^{\sigma}\varphi(\sigma)\,d\sigma, \tag{4.2}$$

with the parameters r_{lm}, α, β satisfying the conditions:

$$r_{lm} = r_{ml},\ \alpha \geqslant 0,\ \beta \geqslant 0,\ \alpha + \beta > 0. \tag{4.3}$$

The derivative of this function which, in accordance with the system (1.4)* must be positive definite for all functions $\varphi(\sigma)$ of the class A, and in particular when

$$\begin{aligned} \frac{1}{2}\frac{dV}{dt} = W(x_l, \sigma) &= \sum_{l=1}^{n}\sum_{m=1}^{n} x_l r_{lm}\left(\sum_{k=1}^{n} a_{mk}x_k + b_m\varphi(\sigma)\right) \\ &+ \alpha\gamma\varphi(\sigma)\left(\sigma - \sum_{l=1}^{n} c_l x_l\right) - \beta\dot{\varphi}(\sigma)\left(\sum_{l=1}^{n} c_l\left(\sum_{k=1}^{n} a_{lk}x_k + b_l\varphi(\sigma)\right) - \gamma\varphi(\sigma)\right) \end{aligned} \tag{4.4}$$

* We note that $\frac{d}{dt}\left(\sigma - \sum_{l=1}^{n} c_l x_l\right) = -\gamma\varphi(\sigma)$.

$$\varphi(\sigma) = h\sigma,\ h>0. \tag{4.5}$$

By putting (4.5) in (4.4) we obtain a quadratic form of real variables x_l and σ which must be positive definite. This implies that for all complex values of x_l and σ not vanishing simultaneously the inequality

$$\operatorname{Re}\left\{\sum_{l=1}^{n}\sum_{m=1}^{n}\overline{x_l}r_{lm}\left(\sum_{k=1}^{n}a_{mk}x_k + b_m h\sigma\right) + \alpha\gamma h\overline{\sigma}\left(\sigma - \sum_{l=1}^{n}c_l x_l\right)\right.$$
$$\left. -\beta h\overline{\sigma}\left(\sum_{l=1}^{n}c_l\left(\sum_{k=1}^{n}a_{lk}x_k + b_l h\sigma\right) - \gamma h\sigma\right)\right\} > 0, \tag{4.6}$$

is valid, the quantities $\overline{x}_l$ and $\overline{\sigma}$ being conjugate to x_l and σ. That this is so can be seen by putting in (4.6) $x_l = u_l + jv_l$, $\sigma = \mu + j\nu$, with u_l, v_l, μ, and ν real; the left-hand side of (4.6) assumes the form $\frac{1}{2}(W_0(u_l, \mu) + W_0(v_l, \nu))$, where W_0 is the quadratic form (4.4) with $\varphi(\sigma) = h\sigma$. Thus, if u_l, v_l, μ, and ν do not vanish simultaneously, we have the inequality (4.6).

The inequality (4.6) is satisfied in the particular case of

$$\sigma = \frac{1}{h},$$
$$x = M_l(j\omega) \qquad (l = 1, 2, \ldots, n), \tag{4.7}$$

where $M_l(j\omega)$ satisfy the system of simultaneous equations

$$j\omega M_l(j\omega) = \sum_{k=1}^{n} a_{lk}M_k(j\omega) + b_l \qquad (l = 1, 2, \ldots, n). \tag{4.8}$$

The system (4.8) has a unique solution for any real ω as according to our assumptions (see section 1) the matrix (a_{lk}) has no purely imaginary eigenvalues.

By taking Fourier transforms of the system (2.1) including (2.2) we obtain

$$j\omega F\{\psi_{lm}(t)\} = \sum_{k=1}^{n} a_{lk}F\{\psi_{km}(t)\} + \delta_{lm} \quad \begin{pmatrix} l = 1, 2, \ldots, n \\ m = 1, 2, \ldots, n \end{pmatrix}, \tag{4.9}$$

where

$$F\{\psi_{lm}(t)\} = \int_0^\infty e^{-j\omega t}\psi_{lm}(t)\,dt. \tag{4.10}$$

By comparing (4.9) and (4.8) we obtain

$$M_l(j\omega) = \sum_{m=1}^{n} F\{\psi_{lm}(t)\}\,b_m = \sum_{m=1}^{n}\int_0^\infty e^{-j\omega t}\psi_{lm}(t)\,b_m dt. \tag{4.11}$$

Substituting (4.7) into (4.6) and using the relations (4.8) we obtain

$$\operatorname{Re}\left\{\sum_{l=1}^{n}\sum_{m=1}^{n}\overline{M_l(j\omega)}\,r_{lm}j\omega M_m(j\omega) + \alpha\gamma\left(\frac{1}{h} - \sum_{l=1}^{n}c_l M_l(j\omega)\right)\right.$$
$$\left. -\beta\left(\sum_{l=1}^{n}c_l j\omega M_l(j\omega) - \gamma\right)\right\} > 0. \tag{4.12}$$

But [see (4.11), (2.6), (2.9)]

$$\sum_{l=1}^{n} c_l M_l(j\omega) = \sum_{l=1}^{n}\sum_{m=1}^{n} \int_0^\infty e^{-j\omega t} c_l \Psi_{lm}(t)\, b_m dt = -\int_0^\infty e^{-j\omega t} v(t)\, dt = -N(j\omega). \tag{4.13}$$

In view of $r_{lm} = r_{ml}$, we have*

$$\operatorname{Re}\sum_{l=1}^{n}\sum_{m=1}^{n} \overline{M_l(j\omega)}\, r_{lm}\, j\omega M_m(j\omega) = \operatorname{Re}\left\{\frac{1}{2} j\omega \sum_{l=1}^{n}\sum_{m=1}^{n} \left(\overline{M_l(j\omega)}\, M_m(j\omega) + M_l(j\omega)\, \overline{M_m(j\omega)}\right) r_{lm}\right\} = 0. \tag{4.14}$$

The inequality (4.12) can therefore be rewritten as

$$\frac{\alpha\gamma}{h} + \operatorname{Re}(\alpha\gamma + j\omega\beta) N(j\omega) + \beta\gamma > 0 \tag{4.15}$$

or [see (2.10)]

$$\frac{\alpha\gamma}{h} + \operatorname{Re}(\alpha\gamma + j\omega\beta) G(j\omega) > 0. \tag{4.16}$$

The above consideration remain valid for any real ω. Thus, it has been shown that for a Lyapunov function of the considered kind to exist it is necessary that the inequality (4.16) be valid for all $h>0$ and for any ω.

If $\alpha \neq 0$, [see (4.3)] and thus if $\alpha\gamma > 0$ [see (1.8)], from (4,16) must follow the inequality

$$\operatorname{Re}\left(1 + j\omega\frac{\beta}{\alpha\gamma}\right) G(j\omega) \geqslant 0. \tag{4.17}$$

Indeed, should the inequality (4.17) be invalid for some $\omega = \omega_0$, then a positive quantity h exists such that the inequality (4.16) will not take place for $\omega = \omega_0$. The necessary condition (4.17) is identical with the inequality (2,11) if $q = \beta/\alpha\gamma > 0$.

If $\alpha = 0$, we obtain from (4,3) and (4.16) the inequalities

$$\beta > 0, \tag{4.18}$$

$$\operatorname{Re} j\omega G(j\omega) > 0. \tag{4.19}$$

We shall now consider certain properties of the function $N(j\omega)$ [see (2.9)]. The function $N(j\omega)$ is a continuous function of ω in view of the inequality (2.8). By the Riemann-Lebesque Lemma

$$\lim_{|\omega|\to\infty} N(j\omega) = 0. \tag{4.20}$$

Therefore, there exists a positive number P_1 such that for any ω

$$\operatorname{Re} N(j\omega) > -P_1. \tag{4.21}$$

By the Riemann-Lebesque Lemma we also obtain** [see (3.3)]

$$\lim_{|\omega|\to\infty} j\omega N(j\omega) = v(0) = -\sum_{l=1}^{n} c_l b_l. \tag{4.22}$$

*We note that $\operatorname{Im}\left(\overline{M_l(j\omega)}\, M_m(j\omega) + M_l(j\omega)\, \overline{M_m(j\omega)}\right) \equiv 0$.

**From the relations (2.6), (2.1) and (2.7) it is easy to find that the integral $\int_0^\infty \left|\frac{dv(t)}{dt}\right| dt$ must converge.

It follows from (4.21), (4.22) and (2.10) that

$$\operatorname{Re} G(j\omega) = \operatorname{Re} N(j\omega) > -P_1, \tag{4.23}$$

$$\lim_{|\omega|\to\infty} j\omega G(j\omega) = \lim_{|\omega|\to\infty} j\omega N(j\omega) + \gamma = -\sum_{l=1}^{n} c_l b_l + \gamma > 0. \tag{4.24}$$

The latter is a necessary condition and can be obtained from (4.6) and (4.18) by putting $x_l = 0$, $\sigma = 1/h$, $\alpha = 0$. It follows from (4.19), (4.24) and from the continuity of the function $G(j\omega)$ that a positive quantity P_2 can be found such that for any real ω the inequality

$$\operatorname{Re} j\omega G(j\omega) > P_2. \tag{4.25}$$

takes place.

By multiplying the inequality (4.25) by $2P_1/P_2$ and by adding the result to the inequality (4.23), we obtain

$$\operatorname{Re}\left(1 + 2\frac{P_1}{P_2} j\omega\right) G(j\omega) > -P_1 + 2P_1 > 0. \tag{4.26}$$

The latter implies that the inequality (2.11) is valid (as a strong inequality) when $q = 2P_1/P_2 > 0$.

Thus it has been proved that the inequality (2.11) suffices for the trivial solution of the system described in section 1 to be absolutely asymptotically stable and also to be a necessary condition for the existenec of a Lyapunov function of the considered type.

Remarks. (1) In order to construct a Lyapunov function of the "quadratic form" type (that is of the kind as in (4.2) with $\beta = 0$) it is necessary that the inequality (4.17) be valid with $\beta = 0$, that is that

$$\operatorname{Re} G(j\omega) \geq 0 \tag{4.27}$$

for any real ω.

The condition (4.27) is sufficient for the absolute asymptotic stability of the trivial solution of the system under investigation, as in this case the inequality (2.11) occurs for $q = 0$.

(2) Most of the Lyapunov functions so far constructed in [8] are of the kind as in (4.2), with $\alpha = 0$. That is why it is necessary that the inequality (4.19) be valid; the latter is also sufficient for the absolute stability of the system trivial solution.

The condition (4.19) can also be written for any real positive ω as

$$\operatorname{Im} G(j\omega) < 0. \tag{4.28}$$

This follows from the relation [see (2.9) and (2.10)]

$$\operatorname{Im} G(j\omega) = -\operatorname{Im} G(-j\omega) \tag{4.29}$$

and from the inequality [see (2.10), (3.7), and (1.8)].

$$[\operatorname{Re} j\omega G(j\omega)]_{\omega=0} = \gamma > 0. \tag{4.30}$$

(3) It would be interesting to find the general solution of the following inverse problem: If the condition (2.11) is satisfied is it always possible to construct a Lyapunov function of the kind as in (4.2)? For some relatively simple cases the answer is in the affirmative.

5. Various Analytic and Graphic Forms of the (2.11) Criterion

The function $(1 + j\omega q)G(j\omega)$ can be written as

$$(1 + j\omega q)\,G(j\omega) = \frac{P(j\omega)}{Q(j\omega)}, \tag{5.1}$$

(see Appendix 1) with P (jω) and Q (jω) being polynomials of jω. The condition (2.11) can now be written as

$$\mathrm{Re}\,P(j\omega)\,Q(-j\omega) \geqslant 0 \tag{5.2}$$

or

$$R(x) \geqslant 0,\ x = \omega^2, \tag{5.3}$$

where R (x) is a polynomial of the variable x. The condition (2.11) is reduced, therefore, to a polynomial of x being nonnegative for $x \geq 0$. The solution can be obtained by using standard algebraic methods.

An arbitrary nonnegative parameter q appears in the criterion, which can be selected in a suitable manner in every specific case.* The algebraic methods for obtaining the optimal values of q are quite straightforward.

The following graphical criteria of absolute asymptotic stability are of special interest in practical applications.

The locus of the points (u, v) in the plane of (u, v) such that

$$u\ (\omega) = \mathrm{Re}\,G(j\omega),\ v(\omega) = \omega \mathrm{Im}\,G(j\omega) \tag{5.4}$$

shall be called the "modified phase-amplitude characteristic" (M.P. A.C.). Directly from (2.11) we obtain the inquality

$$u(\omega) + qv(\omega) \geqslant 0,\ q \geqslant 0, \tag{5.5}$$

which means that the M. P.A.C. is in a half-plane.

A graphical criterion (Fig. 1). If there is a straight line situated either in the first and the third quadrants of the (u, v) plane or it is the ordinate axis**, and in addition it is such that the M.P.A.C. is "on the right" of this straight line, then the trivial solution of the investigated system is absolutely asymptotically stable. One may add that the M.P.A.C. is "on the right" of this straight line if any point of the M.P.A.C. is either on the straight line or it is in the half-plane bounded by this straight line and containing the point $(+\infty, 0)$.***

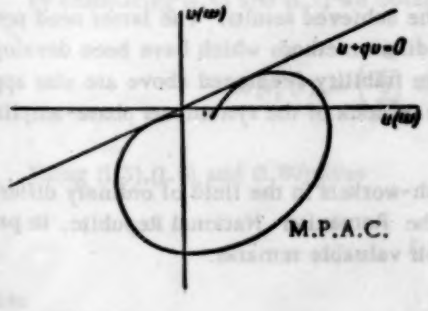


Fig. 1.

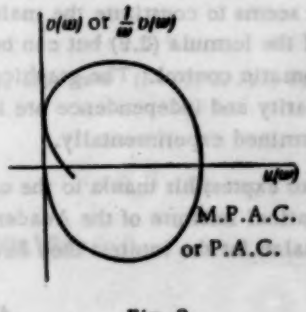


Fig. 2.

* We note that in the criteria (4.27) or (4.28) which described the results obtainable by the usual kind of Lyapunov function (see Remarks (1) and (2), there is no arbitrary parameter.

** Such a straight line obviously passes through the origin. Its equation is of the form $u + qv = 0$ with $q \geq 0$.

*** Or, otherwise, that the inequality (5.5) takes place.

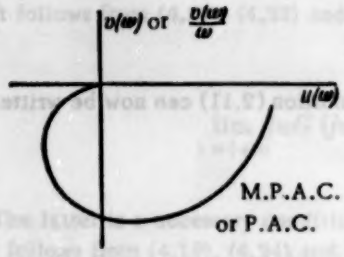


Fig. 3.

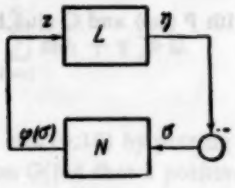


Fig. 4.

The ordinary phase-amplitude characteristic can also be made use of in order to obtain simplified graphical criteria of absolute stability.

The following graphical criterion is obtained from the condition (4.27).

Simplified graphical criterion No. 1. If all the points of the ordinary (or modified) phase-amplitude characteristic are situated "on the right" of the ordinate axis, the trival solution of the system under investigation is asymptotically absolutely stable. (Fig. 2)

It is necessary that this graphical criterion be satisfied in order that a Lyapunov function of the "quadratic form" kind may exist [see Remark (1)].

From the sufficient condition (4.28) of stability another simplified graphical criterion is obtained.

Simplified grapical criterion No. 2. If, when $\omega > 0$, all the points of the ordinary (or modified) P.A.C. are situated in the third or the fourth quadrant or on the negative ordinate semi-axis, then the trivial solution of the system under investigation is asymptotically absolutely stable (Fig. 3).

It is necessary that this criterion be fulfilled in order that a Lyapunov function (5.3) with $\alpha = 0$ may exist [see Remark (2)].

We should like to mention that no simple method exists to express the general graphical criterion (Fig. 1) by means of the ordinary P.A.C. Knowing the ordinary P.A.C., one is able to obtain the modified P.A.C. by multiplying the ordinate of each point by the corresponding value of the variable.

6. Concluding Remarks

The majority of the arguments developed in the preceding sections can be applied with practically no alterations to more general cases mentioned in the introduction, and results of similar nature are obtained.

The fact that in the criteria only the transfer function of the system linear part appears, apart from the simple assumptions of section 1, seems to constitute the main characteristic of the achieved results. The latter need not be evaluated with the aid of the formula (2.9) but can be obtained by more direct methods which have been developed for linear systems of automatic control. The graphical criteria of absolute stability developed above are also applicable when nothing but its linearity and independence are known about the linear block of the system, its phase-amplitude characteristic being determined experimentally.

The author wishes to express his thanks to the collective of research-workers in the field of ordinary differential equations of the Mathematical Institute of the Academy of Sciences of the Romanian National Republic, in particular to Professor A. Khalan, for the interest they have shown and for their valuable remarks.

APPENDIX 1

The investigated system can always be represented in the form of a block diagram as in Fig. 4 where the linear block is denoted by L, the latter described by the system of equations

$$\frac{dx_l}{dt} = \sum_{k=1}^{n} a_{lk} x_k + b_l z \quad (l = 1, 2, \ldots, n),$$

$$\frac{d\xi}{dt} = z,$$

$$\eta = -\sum_{l=1}^{n} c_l x_l + \gamma\xi, \tag{1.1}$$

and N representing the nonlinear block. The input and the output quantities of the linear and nonlinear blocks are related by the equations

$$\eta = -\sigma, \quad z = \varphi(\sigma). \tag{1.2}$$

Let $z^0(t)$ be a known function such that the integral $\int_0^\infty |z^0(t)|$ exists, and let $L\{z^0(t)\}$ be its Laplace transform

$$L\{z^0(t)\} = \int_0^\infty e^{-st} z^0(t)\, dt, \quad \operatorname{Re} s \geq 0. \tag{1.3}$$

Let further $x_l^0(t)$, $\xi^0(t)$, $\eta^0(t)$ be the solution of (1.1) when $z = z^0(t)$ and when the initial conditions are all nul. It follows from the assumptions in section 1 that $L\{x_l^0(t)\}$ exists at least for $\operatorname{Re} s \geq 0$, and $L\{\xi^0(t)\}$ and $L\{\eta^0(t)\}$ exist for $\operatorname{Re} s > 0$.

Taking Laplace transforms of (1.1) we obtain

$$sL\{x_l^0(t)\} = \sum_{k=1}^{n} a_{lk} L\{x_k^0(t)\} + b_l L\{z^0(t)\}, \quad \operatorname{Re} s > 0 \quad (l = 1, 2, \ldots, n). \tag{1.4}$$

$$sL\{\xi^0(t)\} = L\{z^0(t)\}, \quad \operatorname{Re} s > 0, \tag{1.5}$$

$$L\{\eta^0(t)\} = -\sum_{l=1}^{n} c_l L\{x_l^0(t)\} + \gamma L\{\xi^0(t)\}, \quad \operatorname{Re} s > 0. \tag{1.6}$$

The system of simultaneous equations (1.4) has a unique solution when $\operatorname{Re} s > 0$ (see the assumptions in section 1).

Now taking Laplace transforms of the system (2.1) with the initial conditions (2.2):

$$sL\{\psi_{lm}(t)\} = \sum_{k=1}^{n} a_{lk} L\{\psi_{km}(t)\} + \delta_{lm}, \quad \operatorname{Re} s > 0 \quad \begin{pmatrix} l = 1, 2, \ldots, n \\ m = 1, 2, \ldots, n \end{pmatrix}. \tag{1.7}$$

By comparing (1.4) and (1.7) we obtain

$$L\{x_l^0(t)\} = \left(\sum_{m=1}^{n} L\{\psi_{lm}(t)\} b_m\right) L\{z^0(t)\}, \quad \operatorname{Re} s > 0 \quad (l = 1, 2, \ldots, n). \tag{1.8}$$

Using (1.5), (1.6), and (1.80) gives

$$L\{\eta^0(t)\} = \widetilde{G}(s)\, L\{z^0(t)\}, \quad \operatorname{Re} s > 0, \tag{1.9}$$

where

$$\widetilde{G}(s) = N(s) + \frac{\gamma}{s}, \quad \operatorname{Re} s > 0, \tag{1.10}$$

$$N(s) = -\sum_{l=1}^{n} \sum_{m=1}^{n} c_l L\{\psi_{lm}(t)\} b_m, \quad \operatorname{Re} s > 0. \tag{1.11}$$

The function $\widetilde{G}(s)$ defined by the equations (1.10)-(1.11) is the transfer function of the system linear block. The function $\widetilde{G}(s)$ is obtained in the form of a rational function of s. The above considerations are only valid when Re

$s > 0$; we shall say nevertheless that G (s) is a transfer function if it is a rational function defined in the whole s-plane and obtained as the analytic continuation of the $\tilde{G}$ (s) function.

It should be mentioned that the function $L\{\psi_{lm}(t)\}$ exists when Re s = 0 and that when $s = j\omega$ (ω real), then $L\{\psi_{lm}(t)\} = F\{\psi_{lm}(t)\}$ [see (4.10)]. By comparing (1.11) with (2.6) and (2.9) we see that when $s = j\omega$ the function (1.11) is equal to the function of (2.9). Therefore, the function (2.10) is equal to the above defined function G (s) with $s = j\omega$.

APPENDIX 2

By using (3.1) and (3.2) we obtain for $t > T$

$$\lambda_T(t) = -\int_0^T \left(\nu(t-\zeta) + q\frac{d\nu(t-\zeta)}{dt}\right)\varphi(\sigma(\zeta))\,d\zeta. \tag{II.1}$$

But [see (2.6), (2.1) and (2.7)]

$$\left|\nu(t) + q\frac{d\nu(t)}{dt}\right| \leqslant \sum_{l=1}^{n} |c_l|\left(1 + q\sum_{k=1}^{n}\sum_{m=1}^{n}|a_{km}|\right)\sum_{i=1}^{n}|b_i|K_1 e^{-K_0 t}. \tag{II.2}$$

Therefore,

$$|\lambda_T(t)| \leqslant \sup_{0<\zeta<T}|\varphi(\sigma(\zeta))|\int_0^T \left|\nu(t-\zeta) + q\frac{d\nu(t-\zeta)}{dt}\right| d\zeta$$

$$\leqslant \sup_{0<\zeta<T}|\varphi(\sigma(\zeta))|\sum_{l=1}^{n}|c_l|\left(1 + q\sum_{k=1}^{n}\sum_{m=1}^{n}|a_{km}|\right)\sum_{i=1}^{n}|b_i|K_1\int_0^T e^{-K_0(t-\zeta)}\,d\zeta. \tag{II.3}$$

This gives the inequality (3.4), where

$$K_3 = \sup_{0<\zeta<T}|\varphi(\sigma(\zeta))|\sum_{l=1}^{n}|c_l|\left(1 + q\sum_{k=1}^{n}\sum_{m=1}^{n}|a_{km}|\right)\sum_{i=1}^{n}|b_i|\frac{K_1}{K_0}(e^{K_0 T} - 1). \tag{II.4}$$

For a specified solution of the system and for a given T, K_3 is a constant.

Therefore the Fourier transformation (3.5) exists.

APPENDIX 3

From the equation (1.2) we obtain

$$E = \int_0^T \left(\mu_m(t) + q\frac{d\mu_m(t)}{dt}\right)\varphi_T(\sigma(t))\,dt = \int_0^T \left(\mu_m(t) + q\frac{d\mu_m(t)}{dt}\right)\frac{d\xi(t)}{dt}\,dt. \tag{III.1}$$

The integration by parts gives

$$E = \left[\left(\mu_m(t) + q\frac{d\mu_m(t)}{dt}\right)\xi(t)\right]_0^T - \int_0^T \xi(t)\left(\frac{d\mu_m(t)}{dt} + q\frac{d^2\mu_m(t)}{dt^2}\right)dt, \tag{III.2}$$

and hence, in view of $|\xi(T)| \leqslant \sup_{0<\zeta<T}|\xi(\zeta)|$ и $|\xi(0)| \leqslant \sup_{0<\zeta<T}|\xi(\zeta)|$, it follows that

$$E| \leqslant \sup_{0<\zeta<T}|\xi(\zeta)|\left(2\sup_{0<t<T}\left|\mu_m(t) + q\frac{d\mu_m(t)}{dt}\right| + \int_0^T \left|\frac{d\mu_m(t)}{dt} + q\frac{d^2\mu_m(t)}{dt^2}\right| dt\right). \tag{III.3}$$

But [see (2.5), (2.1), and (2.7)]

$$\left|\mu_m(t) + q\frac{d\mu_m(t)}{dt}\right| \leqslant \sum_{l=1}^{n}|c_l|\left(1 + q\sum_{k=1}^{n}\sum_{m=1}^{n}|a_{km}|\right)K_1, \tag{III.4}$$

$$\left|\frac{d\mu_m(t)}{dt} + q\frac{d^2\mu_m(t)}{dt^2}\right| \leqslant \sum_{l=1}^{n}|c_l|\left(1 + q\sum_{k=1}^{n}\sum_{m=1}^{n}|a_{km}|\right)\sum_{k'=1}^{n}\sum_{m'=1}^{n}|a_{k'm'}|K_1 e^{-K_0 t}. \tag{III.5}$$

By substituting (III.4) and (III.5) into (III.3) and by using the inequality

$$\int_0^T e^{-K_0 t}\,dt < \int_0^\infty e^{-K_0 t}\,dt = \frac{1}{K_0}, \tag{III.6}$$

we obtain the inequality (3.26), where

$$\frac{1}{n}K_4 = \sum_{l=1}^{n} |c_l| \left(1 + q \sum_{k=1}^{n}\sum_{m=1}^{n} |a_{km}|\right)\left(2 + \frac{1}{K_0}\sum_{k'=1}^{n}\sum_{m'=1}^{n} |a_{k'm'}|\right) K_1 \tag{III.7}$$

APPENDIX 4

By replacing $\varphi(\sigma(\xi))$ in (2.3) by $d\xi(\zeta)/d\zeta$ [see (1.2)] and by integrating by parts we obtain

$$x_l(t) = \sum_{m=1}^{n} \psi_{lm}(t)\, x_{m0} + \sum_{m=1}^{n} b_m \left[\psi_{lm}(t-\zeta)\,\xi(\zeta)\right]_{\xi=0}^{\zeta=t} - \sum_{m=1}^{n} b_m \int_0^t \xi(\zeta)\,\frac{d\psi_{lm}(t-\zeta)}{d\zeta}\,d\zeta. \tag{IV.1}$$

From (2.1) and (2.7) it follows that $|x_l(t)| \leqslant K_1 n \max(|x_{m0}|)$

$$+ \sup|\xi(\zeta)|\, K_1 \sum_{m=1}^{n} |b_m| \left(2 + \sum_{l=1}^{n}\sum_{k=1}^{n} |a_{lk}| \int_0^t e^{-K_0(t-\zeta)}\,d\zeta\right). \tag{IV.2}$$

As $\int_0^t e^{-K_0(t-\zeta)}\,d\zeta < \int_{-\infty}^{t} e^{-K_0(t-\zeta)}\,d\zeta = \frac{1}{K_0}$, we obtain the inequality (3.35) where

$$K_3 = nK_1,$$

$$K_2 = K_1 \sum_{m=1}^{n} |b_m| \left(2 + \frac{1}{K_0}\sum_{l=1}^{n}\sum_{k=1}^{n} |a_{lk}|\right)$$

APPENDIX 5

The result (3.42) could be obtained by making use of a lemma published by Barbalat [9]. Below we shall give a more direct proof.

We assume for this purpose that (3.42) is not true. Then a positive quantity δ and a sequence of quantities t_i must exist such that

$$|\sigma(t_i)| > \delta, \quad t_i > 0, \quad \lim_{i\to\infty} t_i = +\infty. \tag{V.1}$$

Let the solution in question [see (3.41), (3.36), (3.37), and (1.3)] be

$$\left|\frac{d\sigma(t)}{dt}\right| < M_1, \quad |\sigma(t)| < M_2. \tag{V.2}$$

The sequence t_k can be selected such that

$$t_k - t_{k-1} > \frac{\delta}{M_1}, \quad t_1 > \frac{\delta}{2M_1}. \tag{V.3}$$

When $|t - t_k| < \frac{\delta}{2M_1}$ we have

$$M_2 > |\sigma(t)| = \left|\sigma(t_k) + \int_{t_k}^{t} \frac{d\sigma(t)}{dt}\,dt\right| > |\sigma(t_k)| - M_1|t - t_k| > \frac{1}{2}\delta. \tag{V.4}$$

We also have [see (1.4) and (1.5)]

$$\int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt \geqslant \sum_{k=1}^{N(T)} \int_{t_k - \frac{\delta}{2M_1}}^{t_k + \frac{\delta}{2M_1}} \varphi(\sigma(t))\,\sigma(t)\,dt, \tag{V.5}$$

where N (T) is an index such that $t_{N(T)} + \frac{\delta}{2M_1} \leqslant T$ and $t_{N(T)+1} + \frac{\delta}{2M_1} > T$.

Obviously

$$\lim_{T\to\infty} N(T) = \infty. \tag{V.6}$$

Let $\inf |\varphi(\sigma)| = m$. Then we obtain from (V.5) [see (1.5)]

$$\frac{\delta}{2} < |\sigma| < M_1,$$

$$\int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt \geqslant N(T)\,\frac{\delta}{M_1}\,\frac{\delta}{2}\,m. \tag{V.7}$$

In view of (V.6)

$$\lim_{T\to\infty} \int_0^T \varphi(\sigma(t))\,\sigma(t)\,dt = \infty, \tag{V.8}$$

which contradicts (3.39), and so (3.42) is proved.

APPENDIX 6

In order to prove the relation (3.44) [see (2.3)] it is sufficient, in view of (2.7), to prove that

$$\overline{\lim_{t\to\infty}} \int_0^t \psi_{lm}(t-\zeta)\,\varphi(\sigma(\zeta))\,d\zeta = 0. \tag{VI.1}$$

follows from (3.44).

According to (2.7) we have

$$\left|\int_0^t \psi_{ml}(t-\zeta)\,\varphi(\sigma(\zeta))\,d\zeta\right| \leqslant K_1 \int_0^t e^{-K_0(t-\zeta)}\,|\varphi(\sigma(\zeta))| \tag{VI.2}$$

By applying de l 'Hopital rule (see, for example, [10]) we obtain [see (3.43)]

$$\lim_{t\to\infty} \int_0^t e^{-K_0(t-\zeta)}\,|\varphi(\sigma(\zeta))|\,d\zeta = \lim_{t\to\infty} \frac{\int_0^t e^{K_0\zeta}\,|\varphi(\sigma(\zeta))|\,d\zeta}{e^{K_0 t}} = \frac{1}{K_0}\lim_{t\to\infty} |\varphi(\sigma(t))| = 0, \tag{VI.3}$$

and hence follows (VI.1).

APPENDIX 7

It can easily be seen that the most general negative definite Lyapunov function of the type "a quadratic form plus the integral of the nonlinearity" can be represented as

$$V = \sum_{l=1}^{n}\sum_{m=1}^{n} x_l r_{lm} x_m - \alpha\left(\sigma - \sum_{l=1}^{n} c_l x_l\right)^2 - 2\beta\int_0^{\sigma} \varphi(\sigma)\,d\sigma + 2\sigma\sum_{l=1}^{n} f_l x_l, \tag{VII.1}$$

with r_{lm}, α, β, f_l being constants. The derivative of the V function by virtue of (4.1) assumes the form

$$\frac{1}{2}\frac{dV}{dt} = W(x_l, \sigma) + \sigma\sum_{l=1}^{n} f_l\left(\sum_{k=1}^{n} a_{lk}x_k + b_l\varphi(\sigma)\right)$$

$$+ \sum_{l=1}^{n} f_l x_l\left[\sum_{l=1}^{n} c_l\left(\sum_{k=1}^{n} a_{lk}x_k + b_l\varphi(\sigma)\right) - \gamma\varphi(\sigma)\right], \tag{VII.2}$$

where W (x_l, σ) is given by the expression (4.4).

We shall show that the equalities

$$f_l = 0 \quad (l = 1, 2, \ldots, n) \tag{VIII.3}$$

are necessary. We note that the function $\frac{1}{2}\,dV/dt$ is positive when

$$x_l = x_l^0 = \varepsilon\int_0^{\infty} \psi_{lm}(t)\,dt \quad (l = 1, 2, \ldots, n) \tag{VII.4}$$

$$\varphi(\sigma) = \varepsilon^2\sigma, \tag{VII.5}$$

$$\sigma = 1, \tag{VII.6}$$

where $\underline{m}$ is one of the integers 1, 2, . . . , n; $\psi_{l\,m}(t)$ the functions introduced in section 2, and ϵ an arbitrary (positive or negative) quantity. The function (VII.5) is of course of the class A. By integrating (2.1) we obtain, in view of $\psi_{l\,m}(t)$ [see also (2.2)],

$$-\psi_{lm}(0) = -\delta_{lm} = \sum_{k=1}^{n} a_{lk} \int_0^\infty \psi_{km}(t)\,dt. \tag{VII.7}$$

Consequently [see (VII.4)]

$$\sum_{k=1}^{n} a_{lk} x_k^0 = -\varepsilon\delta_{lm}. \tag{VII.8}$$

It is easily seen that for the values of (VII.4)-(VII.6) the function $\frac{1}{2}\,dV/dt$ can be written as

$$\frac{1}{2}\frac{dV}{dt} = \left(\sum_{l=1}^{n} f_l \sum_{k=1}^{n} a_{lk} x_k^0\right)\varepsilon + O(\varepsilon^2), \tag{VII.9}$$

in which $0(\epsilon^2)$ denotes terms with the property $|\,0(\epsilon^2)\,| < K\epsilon^2$, with K being a constant.

From (VII.9) using (VII.8) we obtain

$$\frac{1}{2}\frac{dV}{dt} = -f_m\varepsilon + O(\varepsilon^2). \tag{VII.10}$$

If $f_m \neq 0$, then for a sufficiently small ϵ the expression (VII.10) has the sign of $-f_m\epsilon$; the latter, however, is arbitrary because the sign of ϵ is also arbitrary. It is therefore necessary that $f_m = 0$ be true. As $\underline{m}$ was taken as arbitrary we see that the equalities (VII.3) must be true.

The function (4.2) is negative when $x_l = 0, \sigma = 1, \varphi(\sigma) = h\sigma, h > 0$. This gives the inequality

$$-\alpha - h\beta < 0 \quad \text{for any} \quad h > 0, \tag{VII.11}$$

which proves finally the relations (4.3).

LITERATURE CITED

1. A. I. Lur'e and E. N. Rozenvasser, Methods of constructing Lyapunov functions of the nonlinear control systems - theory, Proc. of the First International Conference on Automatic Control, Acad. Sci. USSR Press (1961). Sci. Press (1961).
2. V. M. Popov, Criterii de stabilitate pentru sistemele neliniare de reglare automata pe utikizarea transformatei Laplace. Studii si Cercetarii de Energetica. Anul. IX. No. 4, 1959.
3. V. M. Popov, Criterii suficiente de stabilitate asimptotica in mare pentru sistemele neliniare cu mai multe organe oe executie. Studii si Cercetari de Energetica, Anul. IX, No. 4, 1959.
4. V. M. Popov, Nouveaux criteriums pe stabilite por les systemes automatiques non-lineaires. Revue d'Electrotechnique et d'Energetique, vol. V, No. 1, 1960.
5. V. M. Popov, Noi criterii grafice pentru stabilitatea starii stationare a sistemelor automati neliniare. Studii si Cerletari de Energetica. Anul. X, No. 3, 1960.
6. A. M. Letov, The stability of nonlinear systems under control, Gostekhizdat (1951).
7. A. I. Lur'e, Some nonlinear problems in the theory of automatic control, Gostekhizdat (1951).
8. V. A. Yakubovich, Nonlinear differential equations of automatic control systems with a single controlling device, Bull. Leningrad Univ., No. 7 (1960).
9. I. Barbalat, Systemes d'équations différentielles d'oscillatíons non linéaires. Revue de mathematiques pures et appliquées. Acad. R. P. R., vol. 4.,No. 2, 1959.
10. Miron Nicolescu, Analiză Matematică, Vol. II, Bucaresti Editura Tehnică, 1958.

THE CONDITION GOVERNING THE ABSENCE OF OVER-REGULATION IN CERTAIN NONLINEAR AUTOMATIC CONTROL SYSTEMS

G. M. Ostrovskii (Moscow)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 980-985, August, 1961
Original article submitted November 1, 1960

A method is given for finding the conditions governing the absence of over-regulation in certain nonlinear systems.

In investigating control systems, finding the conditions governing the absence of over-regulation in the system when input perturbations of the step type are responded to is of great significance. In [1] we found the conditions governing the absence of over-regulation in certain linear systems. Here we shall study the conditions governing the absence of over-regulation in certain nonlinear systems.

For simplicity we shall first study a second order system:

$$\frac{dx}{dt} = f_1(x, y), \tag{1}$$

$$\frac{dy}{dt} = f_2(x, y). \tag{2}$$

We shall assume that the functions f_1 and f_2 are analytical.

We shall make the transition to a system of equations in x and $\dot{x}$. For this purpose we differentiate Eq. (1):

$$\frac{d^2x}{dt^2} = \frac{\partial f_1}{\partial x} f_1(x, y) + \frac{\partial f_1}{\partial y} f_2(x, y) = F(x, y). \tag{3}$$

Eliminating y from Eq. (1) and (3), we obtain

$$\ddot{x} = \varphi(x, \dot{x}).$$

Making the transition to the coordinates x, $v = \dot{x}$, we obtain the following first order equation:

$$\frac{dv}{dx} = \frac{\varphi(x, v)}{v}. \tag{4}$$

In a sufficiently small neighborhood of the origin the function $\varphi(x, v)$ can be written as [2]

$$\varphi(x, v) = ax + bv + \varphi_1(x, v),$$

where

$$\lim_{\substack{x \to 0 \\ v \to 0}} \frac{\varphi_1(x, v)}{\sqrt{x^2 + v^2}} = 0.$$

Assume that the characteristic equation for the linear approximation of the system of equations (1), (2) has two different negative roots: $-\gamma_1 > -\gamma_2$ which correspond to the two straight lines OA_1 and OA_2 (Fig. 1):

$$v = -\gamma_1 x, \quad v = -\gamma_2 x.$$

These two straight lines have the following properties [2]. All the integral curves may be tangent only to these two straight lines for $x \to 0$, $v \to 0$.

We shall construct sufficiently small isosceles triangles on the straight lines OA_1 and OA_2; the bisectrix of these

triangles coincides with the straight lines. Then if the integral curves enter the triangle OB_1B_2 it follows that they will remain inside the triangle and arrive at the origin while being tangential to the straight line OA_1. On the other hand, if the integral curve enters the triangle OC_1C_2 it will definitely emerge from it through the sides OC_1 or OC_2. From this it is clear that if the integral curve is located inside the triangle OC_1C_2 above the straight line OA_2 it will definitely enter onto the half-plane $x > 0$. We shall now draw the integral curve ODE which passes through the point D (x_0, v_0). This curve can be found by integrating Eq. (4) with the initial conditions $v(x_0) = v_0$ when x varies in the negative direction.

The integral curve that begins above the curve ODE will definitely enter the region $x > 0$. In fact, assume that the integral curve begins at the point G. Then it cannot intersect the curve ODE since the uniqueness theorem would be violated at the point of intersection; it also cannot go to the point O without intersecting the axis $x = 0$, since the integral curve must be tangent to either OA_1 or OA_2 for $x \to 0$. But the integral curve cannot be tangent to the straight line OA_2 in view of what was said above; it also cannot be tangent to OA_1 without intersecting ODE. Therefore, the integral curve must enter into the half-plane $x > 0$; i.e., the system will have over-regulation.

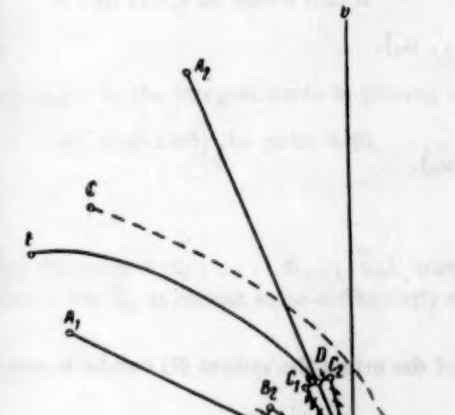


Fig. 1.

Curves beginning below the curve ODE will not make the transition into the half-plane $x > 0$ since they cannot intersect the curve ODE.

Thus the equation for the curve ODE yields the necessary and sufficient condition for the absence of over-regulation in the system.

It should be noted that the reasoning cited above is not sufficiently rigorous. For a rigorous treatment it would be necessary to make the point D on the straight line OA_2 tend toward the origin; then the limiting integral curve would give the condition for the absence of over-regulation. It is easy to see that this integral curve would enter the origin while being tangent to the straight line OA_2. Thus, the curve which yields the condition for over-regulation must be determined by the following conditions:

$$\lim_{x \to 0} \frac{dv}{dx} = -\gamma_2.$$

For example, let us study the equation which describes the system with nonlinear feedback proposed by Lewis:

$$\ddot{x} + (-a|x| + b)\dot{x} + x = 0.$$

It is easy to demonstrate that the curve ODE in this case must intersect the straight line $v = 0$ (Fig. 2). Having found the equation for this curve, it will be possible to find the abscissa for the point of intersection between it and the axis $v = 0$. Then for the initial condition $|x_0| > |\bar{x}_0|$ the system will definitely have over-regulation*.

We shall now study the n-th order system:

$$\frac{dx_1}{dt} = f_1(x_1, \ldots, x_n),$$
$$\ldots\ldots\ldots\ldots$$
$$\frac{dx_n}{dt} = f_n(x_1, \ldots, x_n),$$

where the functions are analytical.

Assume x_1 is the output coordinate of the system and that the conditions governing the absence of over-regulation must be determined for it.

We find the expression for the derivatives of the variable x_1:

$$\frac{dx_1}{dt} = F_1(x_1, \ldots, x_n) = f_1(x_1, \ldots, x_n), \quad \frac{d^2x_1}{dt^2} = F_2(x_1, \ldots, x_n) = \sum_{j=1}^{n} \frac{\partial F_1}{\partial x_j} f_j, \ldots\ldots\ldots$$

* As we know, in a linear system the presence or absence of over-regulation does not depend on the magnitude of the initial deviation.

$$\frac{dx_1^{(n-1)}}{dt^{n-1}} = F_{n-1}(x_1, \ldots, x_n) = \sum_{j=1}^{n} \frac{\partial F_{n-2}}{\partial x_j} f_j, \quad \frac{dx_1^{(n)}}{dt^n} = F_n(x_1, \ldots, x_n) = \sum_{j=1}^{n} \frac{\partial F_{n-1}}{\partial x_j} f_j. \tag{5}$$

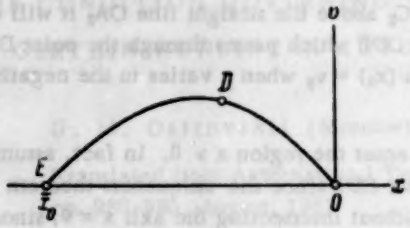


Fig. 2.

We introduce the new variables

$$u_1 = x_1, \; u_2 = \dot{x}_1, \ldots, u_n = x_1^{(n-1)}. \tag{6}$$

From (5) and (6) we obtain the following equations:

$$\begin{aligned} \dot{x}_1 &= u_1, \\ F_1(x_1, \ldots, x_n) &= u_2, \\ &\ldots\ldots\ldots\ldots \\ F_{n-1}(x_1, \ldots, x_n) &= u_n. \end{aligned} \tag{7}$$

Assume that it is possible to solve these equations relative to x_i. We shall substitute the resulting expression for x_i into $F_n(x_1, \ldots, x_n)$. Then we obtain:

$$F_n(x_1, \ldots, x_n) \equiv \Psi(u_1, \ldots, u_n). \tag{8}$$

The system of equations in the new variables will be written as

$$\begin{aligned} \frac{du_n}{dt} &= \Psi(u_1, \ldots, u_n), \\ \frac{du_{n-1}}{dt} &= u_n, \\ &\ldots\ldots\ldots\ldots \\ \frac{du_1}{dt} &= u_2. \end{aligned} \tag{9}$$

We shall study the phase space $u_1, \ldots, u_n$. In the neighborhood of the origin the system (9) can be represented in the form

$$\begin{aligned} \frac{du_n}{dt} &= a_1 u_n + a_2 u_{n-1} + \ldots + a_n u_1 + \psi(u_1, \ldots, u_n), \\ \frac{du_{n-1}}{dt} &= u_n \\ &\ldots\ldots\ldots \\ \frac{du_1}{dt} &= u_2, \end{aligned} \tag{10}$$

where

$$\lim_{u_i \to 0} \frac{\psi(u_1, \ldots, u_n)}{\sqrt{u_1^2 + \ldots + u_n^2}} = 0.$$

We shall assume that the characteristic equation for the linear approximation of the system (10) has n different real negative roots: $-\gamma_1 > -\gamma_2 > \ldots > -\gamma_n$.

In [1] we found the necessary and sufficient condition for the absence of over-regulation in linear systems whose characteristic roots were real and negative. It was demonstrated that if $\Phi(u_1, \ldots, u_n) \leq 0$ for $u_1 < 0$, where

$$\begin{aligned} \Phi = [&\gamma_2\gamma_3 \ldots \gamma_n u_1 + (\gamma_2\gamma_3 \ldots \gamma_{n-1} + \gamma_2\gamma_3 \ldots \gamma_{n-2}\gamma_n + \ldots + \gamma_3\gamma_4 \ldots \gamma_n) u_2 + \\ &\ldots + (\gamma_2\gamma_3 + \ldots + \gamma_2\gamma_n + \ldots + \gamma_{n-1}\gamma_n) u_{n-2} + (\gamma_2 + \gamma_3 + \ldots + \gamma_n) u_{n-1} + u_n], \end{aligned}$$

the integral curve does not enter the half-space $u_1 > 0$.

Just as in [1], we shall study that portion of the phase space which is bounded by the inequalities

$$u_1 < 0, \; u_2 \geq 0, \ldots, u_n \geq 0. \tag{11}$$

From the qualitiative theory of differential equations [2] we know that in the case where the characteristic equation for the linear approximation of the system (10) has different real roots, the arrangement of the integral curves in the neighborhood of the origin for the nonlinear system is essentially similar to the arrangement of the integral

curves for a linear system. We shall now study a sufficiently small region near the origin which is such that we can find the nonlinear system (10) by its linear approximation. Then on the basis of what has been said above there exists a hyperplane

$$A_1u_1 + A_2u_2 + \ldots + A_nu_n = 0 \tag{12}$$

such that if the integral curve is located above this hyperplane then it definitely enters the half-space $u_1 > 0$, i.e., the system has over-regulation. In [1] we demonstrated that this hyperplane consists of the integral curves for the linear system. Of course, the hyperplane (12) can be meaningfully examined only in a sufficiently close vicinity of the origin.

We shall study the hypersurface L:

$$u_n = \varphi\,(u_1, u_2, \ldots, u_{n-1}), \tag{13}$$

which is formed if we continue the integral curves lying in the hyperplane (12) to $t \to -\infty$.

It can easily be shown that if

$$u_{n0} - \varphi\,(u_{10}, \ldots, u_{n-1,0}) > 0$$

and $u_{1,0} < 0$, the integral curve beginning at that point will definitely enter the half-space $u_1 > 0$.

We shall study the point A $(\bar{u}_1, \ldots, \bar{u}_n)$ in n-dimensional space lying on the hypersurface L; i.e.,

$$\bar{u}_n - \varphi\,(\bar{u}_1, \ldots, \bar{u}_{n-1}) \equiv 0,$$

and the point B $(\bar{u}_1, \ldots, \bar{u}_{n-1}, \bar{\bar{u}}_n)$, whose coordinates coincide with the coordinates of the point A up to u_{n-1}, the coordinate $\bar{\bar{u}}_n$ is chosen to be sufficiently cose to $\bar{u}_n$ but in such a manner that

$$\bar{\bar{u}}_n - \varphi\,(\bar{u}_1, \ldots, \bar{u}_{n-1}) > 0$$

The integral curve beginning at the point A cannot enter the half-space $u_1 > 0$ according to the definition of the hypersurface L, since in a sufficiently close vicinity of the origin it will lie in the hyperplane (12).

We shall now study the integral curve emanating from the point B. Based on the continuous dependence of the solutions for the system of differential equations on the initial conditions, it can be stated that if B is sufficiently close to A it follows that for any finite time interval these integral curves will be sufficiently close to one another. The integral curve emanating from B cannot intersect the hypersurface L, since the theorem for the uniqueness of the solution would be violated at the intersection point. In a sufficiently close vicinity of the origin this integral curve will be located "above" the hyperplane (12) and will there definitely enter the half-space $u_1 > 0$. Thus we have demonstrated that if the initial point lies "above" the hypersurface L (no matter how close it lies to this surface), then the integral curve emanating from this point will definitely enter the half-space $u_1 > 0$. If we take the point $\bar{\bar{B}}$, $(\bar{u}_1, \ldots, \bar{u}_{n-1}, \bar{\bar{u}}_n)$ such that $\bar{\bar{u}}_n - \varphi\,(\bar{u}_1, \ldots, u_{n-1}) > 0$, where $\bar{\bar{u}}_n$ differs from $\bar{u}_n$ by any finite amount, it follows from simple mechanical considerations that the integral curve emanating form this point must enter the half-space $u_1 > 0$.

It can also be demonstrated that if $\bar{u}_n - \varphi\,(\bar{u}_1, \ldots, \bar{u}_{n-1}) < 0$, the integral curve does not enter the half-space $u_1 > 0$.

Thus the equation for the hypersurface L yields the necessary and sufficient condition for the absence of over-regulation in the nonlinear system under study. We shall now dwell on the method used for finding the equations defining this hypersurface. It is evident that the hypersurface L consists of integral curves; therefore, if we substitute the equation for the corresponding integral curve into Eq. (13) it will be converted into an identity.

We now differentiate (13) with respect to t:

$$\frac{\partial \varphi}{\partial u_1}\frac{du_1}{dt} + \frac{\partial \varphi}{\partial u_2}\frac{du_2}{dt} + \ldots + \frac{\partial \varphi}{\partial u_{n-1}}\frac{du_{n-1}}{dt} = \frac{du_n}{dt}. \tag{14}$$

Since the functions $u_i(t)$ are components of the integral curve, it follows that their expressions from (9) can be substituted for u_i. Here we find that

$$\frac{\partial \varphi}{\partial u_1} u_2 + \frac{\partial \varphi}{\partial u_2} u_3 + \ldots + \frac{\partial \varphi}{\partial u_{n-1}} u_n = \Psi(u_1, \ldots, u_n). \tag{15}$$

Thus, the equation for the hypersurface L is the solution for an inhomogeneous first order partial differential equation.

For a numerical solution of Eq. (15) there is of course no necessity of seeking an explicit expression for x_i in terms of u_i in order to substitute them into F_n, since for each step of the integration process Eq. (7) can be solved by numerical methods.

Now we need only to determine the boundary condition. It derives from the method by which this surface is constructed. We shall study the region near the origin in which it is possible to treat the linear approximation of the system (10). We shall draw the plane $u_1 = u_{10}$, where u_{10} is sufficiently small. Then the boundary condition can be specified in the form

$$u_1 = u_{10}, u_n = -1/A_n (A_1 u_{10} + A_2 u_2 + \ldots + A_{n-1} u_{n-1}), \tag{16}$$

which is obtained if we substitute $u_1 = u_{10}$ into Eq. (12) and express u_n in terms of the remaining variables.

It is clear that the line of intersection for these planes must lie in the linearity region.

In the analysis above we studied the behavior of integral curves in the half-space $u_1 < 0$. For the half-space $u_1 > 0$ the conditions governing the absence of over-regulation can be formulated as follows: If $u_{n0} - \varphi(u_{10}, \ldots, u_{n-1,0}) \geq 0$, then the integral curve emanating from the point with the coordinates $u_{10}, \ldots, u_{n0}$ does not enter the half-space $u_1 < 0$.

Here we must make the same remark that we made in studying a second order system. Rigorously speaking, the hypersurface L which provides the conditions governing the absence of over-regulation must be tangent to the hyperplane $\Phi(u_1, \ldots, u_n) = 0$ at the origin.

However, the method under study provides a practical means for finding the surface L.

The derived condition governing the absence of over-regulation can be used not only for analyzing an existing control system but also for designing systems in which the characteristics are varied as a function of the position of the image point in the phase space relative to the surface L in order to improve the control performance.

LITERATURE CITED

1. G. M. Osrorvskii, "Increasing the speed of response of certain automatic control systems by means of nonlinear and computing devices", Avtomatika i Telemekhanika, 19, 3 (1958).
2. V. V. Nemytskii and V. V. Stepanov, Qualitative Theory of Differential Equations [in Russian] (State Technical Press, 1948).

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

L. S. PONTRYAGIN'S MAXIMUM PRINCIPLE AND OPTIMAL PROGRAMMING OF ROCKET THRUST•

V. K. Isaev (Moscow)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 986-1001, August,1961
Original article submitted June 10, 1960

The theory of optimal programming of the magnitude and direction of rocket thrust for the cases of plane and space motion in a homogeneous gravitational field is developed, no account being taken of aerodynamic forces. The equations implied by L. S. Pontryagin's maximum principle are integrated in closed form (depending on parameters) for the class of problems considered. The maximum principle is applied to the study of an object described by a system of nonlinear differential equations of fifth (Sections I and II of the paper) and seventh (Section III) orders.

The maximum principle is a generalization of the Weierstrass condition of the classical calculus of variations to the case when the desired optimal control satisfies inequalities of a definite form [1-3]. The maximum principle reduces the variational problem to a boundary problem for a system of ordinary differential equations describing the optimal motion of the object. This system is nonlinear as a rule. To solve it,it is generally necessary to use digital computers.

In this paper the maximum principle is applied to the investigation of one class of problems having to do with the optimal control of the motion of a rocket. The method employed allows one to obtain a broader picture of the nature of optimal processes than do methods based on the classical apparatus. It should be added that questions having to do with the satisfaction of boundary conditions [4,5] are insufficiently treated in papers based on the classical Euler-Lagrange equations.

It is shown on the basis of the maximum principle that for the class of problems considered, the solution may be obtained in the form of a closed parameteric relation among the initial conditions, the optimal control law, and the boundary conditions.

I. MAXIMUM PRINCIPLE

1. Formulation of the Variational Problem

1 . Considering the rocket to be a point of variable mass $m(t) = m(0)\, x_3(t)$, we examine its plane motion in the homogeneous gravitational field of the earth ($g = \text{const}$) in the absence of aerodynamic forces. Suppose $x_1(t)$ and $x_2(t)$ are the horizontal and vertical velocity components, $x_3(t)$ the dimensionless mass, $x_4(t)$ and $x_5(t)$ the range and height, $u_1(t) = P(t)/P_{max}$ the dimensionless thrust (referred to the magnitude of the maximum value P_{max}),and $u_2(t)$ the inclination of the thrust vector to the horizontal (Fig. 1).

Introducing the vectors $x(t) = \{x_1(t), \ldots, x_5(t)\}$ and $u = \{u_1(t), u_2(t)\}$, we write the equations of motion in the form

$$\dot{x}_i = f_i(x, u) \qquad (i = 1, \ldots, 5). \tag{1}$$

By taking components along the earth's axes, and assuming the effective outflow velocity to be constant ($w_l = c$), we obtain the equations:

$$\dot{x}_1 = \frac{A\tilde{u}_1(t)\cos u_2(t)}{x_3(t)}, \qquad \dot{x}_2 = \frac{Au_1 \sin u_2(t)}{x_3(t)} - g,$$
$$\dot{x}_3 = -\alpha u_1(t), \qquad \dot{x}_4 = x_1, \qquad \dot{x}_5 = x_2, \tag{1a}$$

• The material in this paper was presented at the 6th scientific conference at MFTI of April 15, 1960, and at the Moscow General Seminar on nonlinear problems in the theory of automatic control at the IAT AN SSSR in June, 1960.

where $\alpha = P_{max}/m(0)c$, $A = \alpha c$.

The bounds imposed on the magnitude of the thrust•,

$$0 \leq u_1(t) \leq 1, \tag{2}$$

will be basic to our discussion.

2. We pose the following problem: to determine the optimal control (if one exists) $u = \{u_1, u_2\}$, which takes system (1) from a fixed initial position $x(0) = x^0 = \{x_1(0), \ldots, x_5(0)\}$•• in a fixed amount of time ••• $t = T$ to a certain closed convex set G (x), such that at the moment $t = T$ the functional $S = \sum_{i=1}^{n} c_i x_i(T)$ takes on the maximal (minimal) value from among all possible, considering the limitations (2) imposed on the control.

The problem of optimizing an arbitrary differentiable function of a finite number of phase coordinates (Mayer's problem) and of optimizing an integral of an arbitrary function of the coordinates (Lagrange's problem) both reduce to the problem of optimizing a functional such as S above.

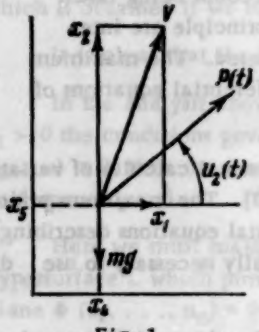


Fig. 1

3. The proof of the maximum principle for the problem with fixed right end of the trajectory was given by L. S. Pontryagin, R. V. Gamkrelidze, and V. G. Boltyanskii [1]. L. I. Rozonér considered boundary conditions for the case when the region G is closed and convex [2] and also for the case of a free right end of the trajectory [3].

In accordance with the procedure of the maximum principle, we supplement system (1) by a conjugate Hamiltonian system:

$$\dot{p}_i = -\sum_{s=1}^{5} p_s \frac{\partial f_s}{\partial x_i} \quad (i = 1, \ldots, 5) \tag{3}$$

or

$$\dot{p}_1 = -p_4, \quad \dot{p}_2 = -p_5, \quad \dot{p}_3 = \frac{Au_1}{x_3^2}(p_1 \cos u_2 + p_2 \sin u_2),$$
$$\dot{p}_4 = 0, \quad \dot{p}_5 = 0. \tag{3a}$$

We consider the fundamental function (the Hamiltonian)

$$H = H(x, p, u) = \sum_{s=1}^{5} p_s f_s(x, u) =$$
$$= u_1\left[\frac{A}{x_3}(p_1 \cos u_2 + p_2 \sin u_2) - \alpha p_3\right] - p_2 g + p_4 x_1 + p_5 x_2. \tag{4}$$

According to the maximum principle, a necessary (and sufficient as well in the case of linear systems [3]) max-optimal•••• (min-optimal) condition is that the control $\underline{u}$ satisfy the minimum (maximum) condition.

For clarity we shall limit ouselves to the consideration of max-optimal control.

2. Minimum Condition

According to the minimum condition the optimal control $u = \{u_1, u_2\}$ is chosen in such a manner that the function H (or, what is the same thing, the function $H^* = u_1\left[\frac{1}{x_3}(p_1 \cos u_2 + p_2 \sin u_2) - \frac{p_3}{c}\right]$, since $\pi(p, x) = H - H^* A$ does not depend on $\underline{u}$) attains an absolute minimum in the set (2) at every moment of time t ($0 \leq t \leq T$) for values of the variables equal to the values of the control at this same moment of time. In the set (2) the function

• The problem may be generalized to the case of arbitrary bounds $\beta_0 \leq u_1(t) \leq 1$ without changing the procedure about to be described.

•• Since $x_3(0)$ is always equal to 1 it is necessary to assume m (0) to be given. This remark should be born in mind whenever the set of initial values x (0) is being described.

••• The practical importance of problems with fixed time was noted by D. E. Okhotsimskii and T. M. Eneev [6] and A. Miele [7].

•••• Max-optimal refers to the control under which the functional S attains a maximum.

$$H^*(x, p, u) = u_1\left[\frac{\sqrt{p_1^2+p_2^2}}{x_3}\sin(u_2+\varphi) - \frac{p_3}{c}\right],$$

where $\sin\varphi = \frac{p_1}{\sqrt{p_1^2+p_2^2}}$, $\cos\varphi = \frac{p_2}{\sqrt{p_1^2+p_2^2}}$,

attains an absolute minimum under the following conditions:

$$u_1 = \operatorname{Sg}\vartheta(p, x), \tag{5}$$

$$u_2 = -\frac{\pi}{2} - \operatorname{arc\,tg}\frac{p_1}{p_2}, \tag{6}$$

where

$$\operatorname{Sg}\vartheta = \begin{cases} 1 & \text{for } \vartheta > 0, \\ 0 & \text{for } \vartheta < 0, \end{cases} \qquad \vartheta(p, x) = \frac{\sqrt{p_1^2+p_2^2}}{x_3} + \frac{p_3}{c}. \tag{7}$$

As L. S. Pontryagin, R. V. Gamkrelidze, and V. G. Boltyanskii have shown[1], maximum principle is equivalent to the Weierstrass condition in the case when U is an open set in the vector space of the variables $u = \{u_1, \ldots, u_r\}$ (in our case the subset u_2, in which, according to condition (6), the control u_2 is chosen, is open), and, on the other hand, the Weierstrass condition generally ceases to be fulfilled at boundary points of the set U. In fact, G. Leitmann, using the Weierstrass condition, obtained an assertion equivalent to (6) for the case of a min-optimal control in the class of problems considered here[8]. However his solution is not complete since it does not contain the basic switching condition (5) which, together with condition (6) follows immediately from the maximum principle.•

Thus, condition (5) evidently may not be obtained on the basis of classical results.

3. Equations of Optimal Motion

Substituting (5) and (6) into (1a) and (3a), we obtain a system of nonlinear differential equations describing the max-optimal motion of system (1):

$$\dot{x}_1 = -\frac{A\operatorname{Sg}\vartheta}{x_3}\frac{p_1}{\sqrt{p_1^2+p_2^2}}, \quad \dot{x}_2 = -\frac{A\operatorname{Sg}\vartheta}{x_3}\frac{p_2}{\sqrt{p_1^2+p_2^2}} - g, \quad \dot{x}_3 = -\alpha\operatorname{Sg}\vartheta,$$

$$\dot{x}_4 = x_1, \quad \dot{x}_5 = x_2, \quad \dot{p}_1 = -p_4, \quad \dot{p}_2 = -p_5, \quad \dot{p}_3 = -\frac{A\operatorname{Sg}\vartheta}{x_3^2}\sqrt{p_1^2+p_2^2}, \quad p_4 = 0, \quad p_5 = 0 \tag{8}$$

The necessary boundary conditions, according to L. I. Rozonoér[2], may be written in the form

$$p_i(T) = -c_i - \mu b_i[x_1(T), \ldots, x_5(T)], \quad (i = 1, \ldots, 5), \quad F[x_1(T), \ldots, x_5(T)] = 0, \tag{9}$$

where the inequality $F(x_1, \ldots, x_5) \leq 0$ defines the set G, and

$$b_i = \frac{\partial F(x_1, \ldots, x_5)}{\partial x_i} \quad (i = 1, \ldots, 5)$$

for the case when the function $F(x_1, \ldots, x_5)$ is differentiable.

In the other case, which is important in practice, when for $t = T$, q ($q << 4$) coordinates are fixed, i.e., $x_j(T) = x_j^1$ ($j = 1, \ldots, q$), conditions (9) are replaced by the set of conditions:

$$x_j(T) = x_j^1 \quad (j = 1, \ldots, q), \quad p_k(T) = -c_k \quad (k = q+1, \ldots, 5). \tag{10}$$

The six conditions (9), containing one unknown constant $x_i(0) = x_i^0$ ($i = 1, \ldots, 5$), form a set of ten boundary conditions for the solution of system (8) of ten equations in ten unknown $x_i(t)$, $p_i(t)$, ($i = 1, \ldots, 5$).

• According to [8],

$$u_1 = 1 \text{ for } \vartheta > 0, \quad u_1 = 0 \text{ for } \vartheta < 0. \tag{5•}$$

From the point of view of the maximum principle, the condition $\vartheta = 0$ corresponds to the regime of singular control (see Appendix I). There is an obvious class of problems for which (5) and (5•) lead to different qualitative representation.

4. Certain Particular Cases

We briefly consider individual examples, based on conditions (9) and (10).

1. Suppose that after a time interval t = T a rocket with starting weight m(0)g is required to attain a height of $x_5(T) = x_5^1$ with a useful load $m(T)g = m(0)gx_3(T)$ in such a manner that the velocity vector at t = T has an angle $\gamma = \arctan k$, $k \neq 0$ with the horizontal, and the range $x_4(T)$ assumes the maximum of all possible values.

Then

$$c_4 = 1,\ p_4(t) \equiv p_4(T) = -c_4 = -1,\ p_1 = 1. \tag{11}$$

Setting in condition (9) $c_1 = c_2 = 0$, we obtain

$$p_1(T) + kp_2(T) = 0. \tag{12}$$

Using the sixth, seventh, ninth, and tenth equations of system (8) and equations (11) and (12), we find

$$p_2 = -p_5 \equiv -p_5^0 = -\frac{1}{k}\left(\frac{p_1^0 + kp_2^0}{T} + 1\right). \tag{13}$$

As a result we obtain a system of eight equations: the first through the fifth and the eighth equations of system (8) and equations (11) and (13), with eight unknowns $x_i(t)$ $(i = 1, \ldots, 5)$, p_1, p_2, p_3, and eight boundary conditions:

$$x_i(0) = x_i^0 \quad (i = 1, \ldots, 5), \qquad x_2(T) - kx_1(T) = 0, \qquad x_3(T) = x_3^1,$$
$$x_5(T) = x_5^1. \tag{14}$$

2. Suppose in addition to the conditions of the preceeding problem we have k = 0. The optimal process is described by a system of nine equations: the first through the fifth and the seventh, eight, and ninth equations of system (8) and equation (11) in the nine unknowns $x_i(t)$ $(i = 1, \ldots, 5)$, $p_1(t)$, $p_2(t)$, $p_3(t)$, $p_5(t)$, and the nine boundary conditions:

$$x_i(0) = x_0^i \quad (i = 1, \ldots, 5), \qquad x_2(T) = 0, \qquad x_3(T) = x_3^1, \qquad x_5(T) = x_5^1,$$
$$p_1(T) = 0. \tag{15}$$

Thus, if in the general problem of deriving $(k \neq 0)$ $p_5^0 = p_5^0(p_1^0, p_2^0, k, T)$, p_1^0 and p_2^0 are unknown (not counting p_3^0), then for k = 0, p_2^0 and p_5^0 are subject to definition, and $p_1^0 = -T$.

3. The problem of attaining a maximal height at fixed range reduces to a system of equations similar (with other conditions equal) to those described in 1 and 2, under the condition $p_5(T) = -1$.

4. In the problem of finding the ballistic rocket of maximal range (with time of flight T known), the velocity vector $V(T) = \{x_1(T), x_2(T)\}$ is free and the boundary conditions take the following form:

$$x_i(0) = x_i^0 \quad (i = 1, \ldots, 5), \qquad x_3(T) = x_3^1, \qquad x_5(T) = 0,$$
$$p_1(T) = 0, \qquad p_2(T) = 0, \qquad p_4(T) = -1. \tag{16}$$

II. ANALYSIS OF THE COMPOSITION OF THE OPTIMAL TRAJECTORY

We draw attention to the fact that the sixth, seventh, ninth, and tenth equations of system (8) do not depend on $\underline{x}$ and may be integrated separately. Moreover, the functions $p_1(t)$ and $p_2(t)$ are linear with respect to time, by virtue of which the optimal trajectory of image points in the p_1, p_2 plane (the p-trajectory) is a straight line segment

$$-p_4^0 p_2 + p_5^0 p_1 = -p_4^0 p_2^0 + p_5^0 p_1^0 = C, \tag{17}$$

depending parametrically on $t\ (0 \leq t \leq T)$.

We introduce into the investigation a vector space $P\{p_1, p_2, z = -p_3/c\}$ (Fig. 2), in which conditions (5) and (6) may take on a simple geometrical meaning.

1. Direction of Thrust

The program of variation in the angle of inclination of the thrust vector with respect to the horizontal is determined by choosing a p-trajectory from the set of all those admissible; namely, the angle $u_2(t)$ is included between

the axis $-p_1$ and the vector $R(p^0, t) = \{p_1(t), p_2(t)\}$ (the first quadrant is between the negative semiaxes $-p_1$ and $-p_2$).

We call a p-trajectory passing through the origin ($C = 0$) singular since on those for which $p_1^0 < 0 < p_1^1$, there is a discontinuity of the control $u_2(t)$ at the moment $t^1 = p_1^0/p_4^0$:

$$u_2(t^1 + 0) = u_2(t^1 - 0) + \pi. \tag{18}$$

In the general case of optimization of an arbitrary functional S, the tangent of the angle of inclination of the axis of the moving object with respect to the horizontal is a linear-fractional function of time

$$\operatorname{tg} u_2(t) = \frac{p_2^0 - p_5^0 t}{p_1^0 - p_4^0 t}, \tag{19}$$

which corresponds to the result obtained by D. Lowden [9, 10] and G. Leitmann [8].

For the cases when no limitation in range is imposed (the fourth equation of (8) is excluded), we have $p_1 = 0$ and, consequently,

$$\operatorname{tg} u_2(t) = \frac{p_2^0 - p_5^0 t}{p_1^0}, \tag{20}$$

which contains the solutions obtained by D. E. Okhotsimskii and T. M. Éneev [6], and also those by S. Ross [11] for the case $S = x_1(t)$ for $k = 0$, and by B. Fried for the case $S = x_5$ [12].

In problem 4 (sec. 4), using the eighth and the ninth condition (16), we obtain

$$\operatorname{tg} u_2(t) = \frac{p_2}{p_1} = \frac{-p_5^0 t + p_5^0 T}{t - T} = -p_5^0 = \text{const}, \tag{21}$$

i.e., the program of reducing a ballistic rocket of maximal range (fixing the flight time T) in the homogeneous gravitational field of the earth, the angle of inclination of the thrust vector is characterized as constant on a portion of the trajectory which coincides with the solutions of B. Fried and J. Richardson [13], and A. Miele and J. Capellari [5]•.

2. Magnitude of the Thrust

We write the switching function in the following form:

$$\Phi(p, x) = \frac{\sqrt{p_1^2 + p_2^2}}{x_3(t)} + \frac{p_3}{c} = \frac{R(p^0, t)}{x_3} - Z(p, x) = \Phi(p, x) - Z(p, x). \tag{2.2}$$

For fixed values of p^0, $\Phi(p, x)$ and $Z(p, x)$ will be functions of t: $\Phi(t)$ and $Z(t)$.

The active portion of the optimal trajectory in the space X ($t_1^{(l)} < t < t_2^{(l)}$) corresponds in the ω plane (passed through the p-trajectory perpendicularly to the p_1p_2 plane) to the condition $\Phi(t) > Z(t)$, in which $Z(t)$ is a monotonic increasing function on the active portion and is functionally constant on the passive portion [see the eighth equation of (8)] $\tau_1^{(k)} < t < \tau_2^{(k)}$, and where $\Phi(t)$ is described at the same time by the equation of a hyperbola $\Phi(t) = R(p^0, t)/x_3(\tau_1^{(k)})$, $(\tau_1^{(k)} < t < \tau_2^{(k)})$, $t_1^{(l)}(\tau_1^{(k)})$ and $t_2^{(l)}(\tau_2^{(k)})$ are the moments of beginning and ending of the active (passive) portion with index $l(k)$.

Using the space P and the switching condition (5), one may establish many properties of the optimal control, which form the basis of the following assertions, covering the known results obtained with the help of classical methods [5, 8].

Theorem 1. For the class of problems considered (1 and 2, sec. 1) all controls u_1 which are max-optimal with respect to S and the max-optimal trajectories x in the space X corresponding to them, and proceeding from the point x^0 at the time $t = 0$, are contained in the set of controls (and trajectories corresponding to them) satisfying the following properties:

1) The optimal control is of boundary type; in the composition of the optimal control there enter not more than two active portions, $l \leq 2$, the control on which is maintained in the regime of maximal thrust $u_1(t) \equiv 1$ $(t_1^{(l)} < t < t_2^{(l)})$;

• Equation (21) is valid for any functional S under the unique condition that the velocity $V(T) = \{x_1(T)\ x_2(T)\}$ be free.

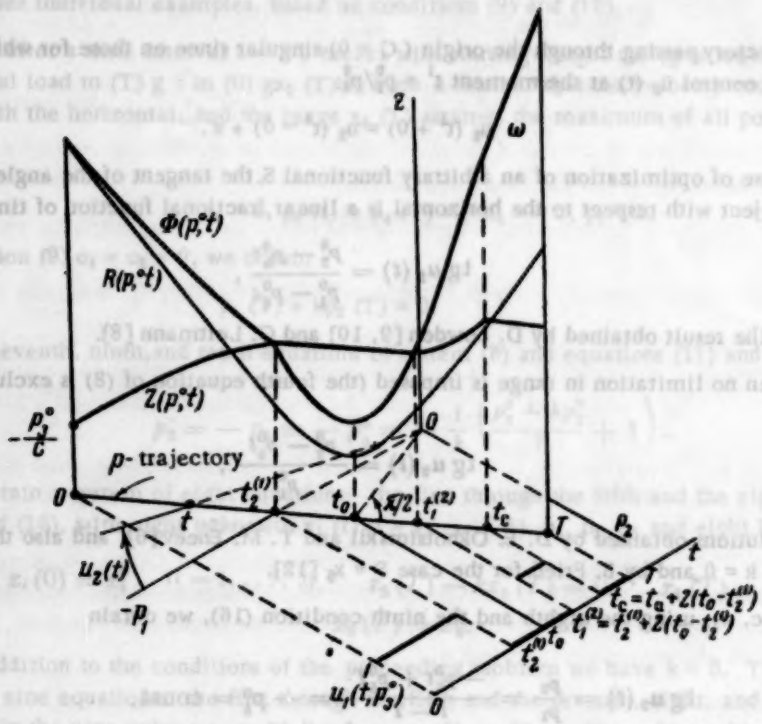


Fig. 2.

2) when there are two active portions present in the composition of the control: a) $t_1^{(1)} = 0$, b) the switching time $t_2^{(1)}$ and $t_1^{(2)}$ of the optimal control u_1 corresponding to some p-trajectory are symmetric with respect to the points $p_1(t_0)$, $p_2(t_0)$ lying on the given p-trajectory and occurring at minimal distance from the origin.

The proof of this theorem is given in Appendix I; it is illustrated in Fig. 2, constructed for some p-trajectory of problem 1 (sec. 4), where the t-axis is collinear with the p_1 axis. In Fig. 2 a particular case is presented, when the point $t_2^{(1)}$ lies on the intersection of the p-trajectory and the $-p_1$ axis. The motor cut-off time t_c is determined not by condition (5), but from the boundary condition $x_3(T) = x_3^1$, i.e., from the condition that the entire fuel load be burned up; $t_2^{(2)} = t_a + 2(t_0 - t_2^{(1)})$, where $t_a = 1 - x_3^1/\alpha$, is the general continuability of the operation of the motor in the regime of maximal thrust•.

3. On Integration of the Equations of Optimal Motion (8)

The entire max-optimal control u_1 is determined by the prescription of the parameters of the p-trajectory and the parameter $z^0 = -p_3^0/c$. On the basis of Theorem 1, one may exclude from consideration the parameter p_3^0, and determine some control u_1 from among those which are optimal [satisfy condition (5)], existing (see Appendix II) on the given p-trajectory [for the given $u_2(t)$]. In this the control with two active portions may be characterized by the parameter $\eta \equiv t_1^{(2)} - t_0 = t_0 - t_2^{(1)}$, defined inside the closed region $Q_{p\eta}$: $(t_0 - t_a)\,\mathrm{Sg}\,(t_0 - t_a) \le \eta \le \min(T - t_a/2, t_0)$, and the controls with one active portion by the parameter $\xi \equiv t_1^{(1)}$, defined on the set of p-trajectories Ω_1 (see Appendix II) in the region $Q_{p\xi}^{(1)}$: $0 \le \xi \le T - t_a$, on the set Ω_2 in the region $Q_{p\xi}^{(2)}$: $2t_0 \le \xi^{(2)} \le T - t_a$, on Ω_3 and Ω_4 $\xi^{(3)} = \xi^{(4)} \equiv 0$.

Thus for the class of problems considered, the entire max-optimal control $u = \{u_1, u_2\}$ is established with the aid of finite operations i.e., by giving the parameters of the p-trajectory and the parameter $\xi(\eta)$, without integrating the system (8). Furthermore, for each control, from the number of optimal equations (8), there is an equation with known right side, integrable by parts into elemntary functions. To separate out the desired control from the set of all optimal controls, the boundary conditions are used.

We shall clarify all this by considering problems 1, 2 (Sec. 4) as examples. In the given case, two out of three

• In problems where the fuel expenditure is minimized, the cut-off moment t_c is determined from the boundary condition $Z(t_c) = 1/c$.

parameters, determining $u_2(t)$ in the general case, are unknown: p_1^0 and p_2^0 (1) or p_2^0 and p_3^0 (2). Consequently, the vector $u(t) = \{u_1, u_2\}$ depends on three parameters: $p_1^0, p_2^0, \xi(\eta)$ (1) or $p_2^0, p_3^0, \xi(\eta)$ (2). The parameters ξ and η together play the role of the single parameter ξ, measured for convenience from the various zero positions on each of the nonintersecting regions $Q_{p\xi}$ and $Q_{p\eta}$ of the p-trajectory.

Integrating (8), we obtain (see Appendix III):

$$x_i(t) = F_i(p_1^0, p_2^0, \zeta, t) + X_i^0 \qquad (i = 1, \ldots, 5), \tag{23}$$

where X_i^0 are the components depending on the initial conditions and independent of the parameters. To determine p_1^{0*}, p_2^{0*}, and ξ^* we use (14):

$$\begin{aligned} F_2(p_1^{0*}, p_2^{0*}, \zeta^*, T) - kF_1(p_1^{0*}, p_2^{0*}, \zeta^*, T) &= kx_1^0 - x_2^0 + gT, \\ F_3(p_1^{0*}, p_2^{0*}, \zeta^*, T) + 1 &= x_3(T), \\ F_5(p_1^{0*}, p_2^{0*}, \zeta^*, T) + x_2^0 T - \frac{gT^2}{2} + x_5^0 &= x_5(T). \end{aligned} \tag{24}$$

Finding the solution to system (24) $(p_1^{0*}, p_2^{0*}, \xi^*)$ with the aid of iteration and setting it into (23) with $i = 4$, we obtain the final result,

$$x_4(T) = F_4(p_1^{0*}, p_2^{0*}, \zeta^*, T) + x_4^0 = \max. \tag{25}$$

The following should also be born in mind: Considering the control $u = u(p_1^0, p_2^0, \xi)$ and the trajectory (23) corresponding to it as a finite parametric solution of the problem for initial conditions $x_1^0, x_3^0, x_4^0, x_5^0$, one may consider the first equation of system (24) in the nature of an equation for $x_2(0) = x_2^{0*}$, and the third as a relation for obtaining $x_5(t)$ with account taken of the value of x_2^{0*} found. The second equation of system (24) serves as a relation for the choice of the motor cut-off point when the fuel has run out.

III. ON THE OPTIMAL MOTION OF A ROCKET IN SPACE

The theory of optimal programming of the rocket thrust presented above may be generalized to the case of spacial motion without difficulty.

If $y_1(t)$, $y_2(t)$, $y_3(t)$ are the velocity components along the earth's axes $x = y_5$, $y = y_6$, and $z = y_7$, $u_2(t)$ and $u_3(t)$ give the direction of the thrust as shown in Fig. 3, $y_4 = m(t)/m(o)$, and $u_1(t) = p(t)/p_{max}$ is the magnitude of the dimensionless thrust, then the equations of motion have the form:

$$\begin{aligned} \dot{y}_1 &= \frac{Au_1 \cos u_2 \cos u_3}{y_4}, \qquad \dot{y}_2 = \frac{Au_1 \cos u_2 \sin u_3}{y_4}, \\ \dot{y}_3 &= \frac{Au_1 \sin u_2}{y_4} - g, \qquad \dot{y}_4 = -\alpha u_1, \qquad \dot{x}_5 = x_1, \\ \dot{y} &= y_2, \qquad \dot{y}_7 = y_3. \end{aligned} \tag{26}$$

Suppose the variational problem considered for system (26) is a generalization of problem 2 (Sec. 1, I) to the case of space motion.

Having limited ourselves to the analysis of max-optimal control, we obtain conditions which the optimal control should satisfy, in accordance with the maximum principle:

$$\begin{aligned} u_1(t) &= \operatorname{Sg} \theta = \operatorname{Sg}\left[\frac{\rho}{y_4(t)} + \frac{\psi_4(t)}{c}\right], \\ \sin u_2(t) &= -\frac{\psi_2(t)}{\rho(t)}, \qquad \cos u_2(t) = \frac{r(t)}{\rho(t)}, \\ \sin u_3(t) &= -\frac{\psi_2(t)}{r(t)}, \qquad \cos u_3(t) = -\frac{\psi_1(t)}{r(t)}, \end{aligned} \tag{27}$$

where

$$\rho(t) = \sqrt{\psi_1^2(t) + \psi_2^2(t) + \psi_3(t)}, \qquad r(t) = \sqrt{\psi_1^2(t) + \psi_2^2(t)}, \tag{28}$$

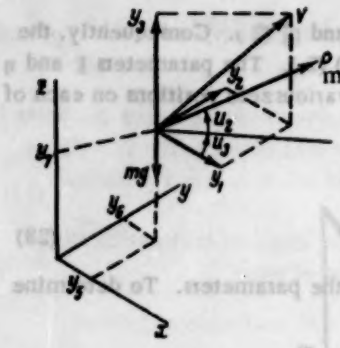


Fig. 3.

in which the vector (t) = $\{ y_1 (t), \ldots, y_7 (t)\}$ and the vector ψ (t) = $\{\psi_1 (t), \ldots, \psi_7 (t)\}$ satisfy the following Hamiltonian system of equations:

$$\dot{y}_1 = -\frac{\psi_1 A \operatorname{Sg}\theta}{y_4 \rho(t)}, \quad \dot{y}_2 = \frac{\psi_2 A \operatorname{Sg}\theta}{y_4 \rho(t)}, \quad \dot{y}_3 = -\frac{\psi_3 A \operatorname{Sg}\theta}{y_4 \rho(t)} - g,$$

$$\dot{y}_4 = -\alpha \operatorname{Sg}\theta, \quad \dot{y}_5 = y_1, \quad \dot{y}_6 = y_2, \quad \dot{y}_7 = y_3, \quad \dot{\psi}_1 = -\psi_5, \quad \dot{\psi}_2 = -\psi_6$$

$$\dot{\psi}_3 = -\psi_7, \quad \dot{\psi}_4 = -\frac{p(t) A \operatorname{Sg}\theta}{y_4^2}, \quad \dot{\psi}_5 = 0, \quad \dot{\psi}_6 = 0, \quad \dot{\psi}_7 = 0, \quad (29)$$

and also the corresponding set of initial and boundary conditions, obtained for the given set G(y) and the functional S with the help of the procedure described above (sec. 3, I).

In proving Theorem 1 for the plane case (see Appendix I), the following properties of the function $\vartheta(p, x) = \frac{R(p^0, t)}{x_3} - Z(p^0, t)$, where $R(p^0, t) = \sqrt{p_1^2(t) + p_2^2(t)}$ were used: a) p_1 (t) and p_2 (t) are linear functions of time (consequently, R (p^0, t) is a hyperbolic function of time); b) Z (p^0, t) is a monotonic increasing function of time on the active portion ($\vartheta > 0$) and a constant function on the passive part ($\vartheta < 0$); and c) the function ϑ is continuous everywhere, on an arbitrary p-trajectory, and its derivative $\dot{\vartheta} \equiv \frac{\dot{R}(p^0, t)}{x_3 t}$ is continuous everywhere on an arbitrary nonsingular p-trajectory and suffers a discontinuity only at the origin ($p_1 = 0$, $p_2 = 0$).

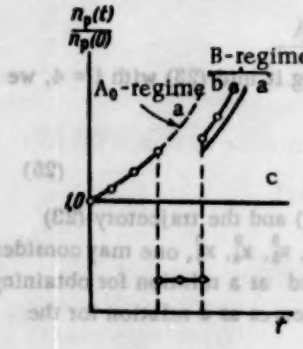


Fig. 4.

By vitrue of the complete analogy between systems (29) and (8) and the analytical properties of the switching functions θ and ϑ, all results of Theorem 1 may be transferred to the case of spacial motion in a homogeneous gravitational field with no account taken of aerodynamical forces.

Consequently, for the general case of motion in space, the problem of the absence of portions with variable thrust in the optimal trajectory (posed by A. Miele and H. Capellari [5] and analyzed by G. Leitmann [8] for plane case) is solved.

Thus, the reactive overload $n_p(t) = \frac{P_{\text{макс}} u_1(t)}{m(0) x_3(t) g}$ on the active portion may only increase, as shown in Fig. 4 a,b (for the case $\beta_0 \leq u_1(t) \leq 1$) and not remain constant, as was found by Yu. A. Gorelov [4] for the particular cases of motion of a rocket with control u_1 chosen from the class C_1 of continuous functions (Fig. 4 c).

The author thanks L. I. Rozonoér for discussing the paper and for many important comments.

APPENDIX I

Proof of Theorem 1

A. On an arbitrary p-trajectory the switching function ϑ (p, x) is continuous and has a continuous derivative with respect to time almost everywhere •. In fact, on the active part, considering the third and eighth equation of system (8), we have:

$$\dot{\Phi}_a = \frac{\dot{R}(p^0, t)}{x_3(t)} + \frac{\alpha R(p^0, t) u_1(t)}{x_3^2(t)}, \quad \dot{Z}_a = \frac{\alpha R(p^0, t) u_1(t)}{x_3^2(t)}. \quad (1.1)$$

On the passive portion:

$$\dot{\Phi}_n = \frac{\dot{R}(p^0, t)}{x_3(t)}, \quad \dot{Z}_n = 0. \quad (1.2)$$

Thus, everywhere on an arbitrary p-trajectory $\dot{\vartheta}(p^0, t) \equiv \frac{\dot{R}(p^0, t)}{x_3(t)}$ $(0 \leqslant t \leqslant T)$. Since $x_3(t) > 0$ $(0 \leq t \leq T)$ is a bounded function, $\vartheta(p^0, t)$ is a monotonic function of differing sign in the regions $0 \leq t \leq t_0$ (A) and $t_0 < t \leq T$ (B), vanishing (if one excludes the case of singular trajectories) at the point $t = t_0$, at which $\dot{R}(p^0, t_0) = 0$. Thus, on the basis of Rolle's theorem, if $t_0 \in [0, T]$, for $0 \leq t \leq T$, the function ϑ (p^0, t) may vanish at no more than

• On singular trajectories at the origin ($p_1 = p_2 = 0$) ϑ undergoes a discontinuity. This circumstance does not influence the coune of the proof, and therefore without violating any generality, we limit ourselves to the consideration of nonsingular trajectories.

two points: $t = t_2^{(1)} < t_0$ and $t = t_1^{(2)} > t_0$. The points $t_2^{(1)}$ and $t_1^{(2)}$ correspond to the ending time of the first and the beginning of the second active portions, since for $t < t_2^{(1)} \in A, \dot{R}(p^0, t) < 0$, i.e., $\vartheta(p^0, t) > 0$*, and for $t > t_1^{(2)} \in B, \dot{R}(p^0, t) > 0$, i.e., $\vartheta(p^0, t) > 0$ again. The function $\vartheta(p^0, t)$ may vanish if $t_0 \in [0, T]$ at not more than one point $t = t_c$, which for p-trajectories situated in the region A, is the ending time for the active portion $t_c = t_2^{(1)}$, and for p-trajectories in the region B, the beginning point of the active portion $t_c = t_1^{(1)}$, which follows from consideration of the switching condition (5) and relations (I.1) and (I.2).

We note that the presence of two zeros of the function $\vartheta(p^0, t)$ is necessary but not sufficient for the realization of two active portions, since besides condition (5), it is necessary to consider boundary conditions. In Fig. 5b the case 4 illustrates these positions for the problem with fixed finite mass.

In order to complete the proof of sec. A, it remains to be shown than in the class of problems considered, the case of singular control is absent. Actually, to realize a singular control on the finite segment $t_1 < t < t_2$ (regime with variable thrust), L. I. Rozonoér [2] has shown that it is necessary that the function H be independent of u everywhere in $t_1 < t < t_2$; i.e., in our case, the condition $\vartheta(p^0, t) \equiv 0$ $(t_1 < t < t_2)$, be fulfilled. By virtue of the continuity of the function $\vartheta(p^0, t)$, the singular control may be conjugate to a nonsingular one at the points t_1 and t_2, where

$$\Phi(p^0, t_i) = Z(p^0, t_i) \qquad (i = 1, 2). \tag{I.3}$$

Besides (I.3), for the existence of a finite portion of a trajectory with singular control, it is necessary that the condition

$$\dot{\Phi}(p^0, t) = \dot{Z}(p^0, t) \qquad (t_1 < t < t_2). \tag{I.4}$$

be fulfilled.

But (I.4), accounting for (I.1) and (I.2), leads to the identity

$$\dot{R}(p^0, t) \equiv 0 \qquad \text{for} \qquad t_1 < t < t_2, \tag{I.5}$$

which in general is not true on singular trajectories (C = 0), and on all remaining, is fulfilled only at the point $t = t_0$.

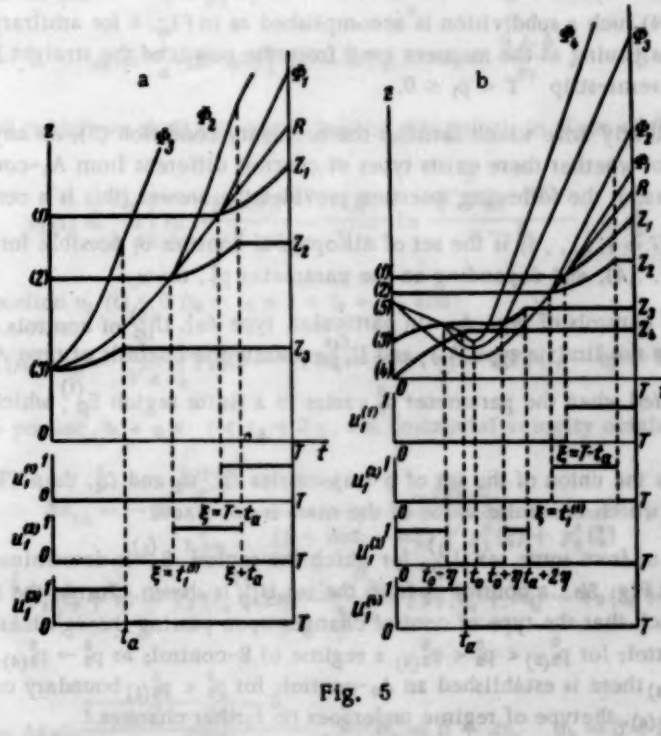


Fig. 5

Thus, for the class of problems considered, the control u_1 remains always on the boundaries of region (2), which was to be proven.

* Which proves assertion 2a of Theorem 1, since for $t < t_2^{(1)} \in A, \vartheta(p^0, t) > 0$.

B. Suppose the function $\vartheta(p^0, t)$ possesses two zeros: $t = t_2^{(1)}$ and $t_1^{(2)}$. Then for $t_2^{(1)} < t < t_1^{(2)}$, $u_1(t) = 0$ (see preceding divisions of the proof) and the function $\Phi(p^0, t) = R(p^0, t)/x_3(t) = R(p^0, t)/x_3(t_2^{(1)})$ is described by the equation of a hyperbola, in which $\dot{\Phi}(p_0, t_0) = \dot{R}(p^0, t_0) = 0$, and the straight line $z(t) = Z(p^0, t) = Z(p^0, t_2^{(1)})$ $(t_2^{(1)} < t < t_1^{(2)})$, parallel to the p_1, p_2 plane [see the eighth equation of system (8)], is a chord of the hyperbola

$$z(t) = \frac{\sqrt{(-p_4^0 t + p_1^0)^2 + (-p_5^0 t + p_2^0)^2}}{x_3(t_2^{(1)})}.$$

It is not necessary to apply rotation of axes to ransform this hyperbola to canonical form. From this it immediately follows that $t_1^{(2)} - t_0 = t_0 - t_2^{(1)}$, where t_0 is determined from the relation

$$t_0 = \frac{p_1^0 p_4^0 + p_2^0 p_5^0}{p_4^{02} + p_5^{02}}, \tag{1.6}$$

which was to be proved.

APPENDIX II

Types of Control Regimes u_1 and Geometry of p-Trajectories

We introduce the following definition: A control of type A_0 (an A_0-control) is a control u_1 with one active portion ($L = 1$) in which $t_1^{(1)} = 0$; a control of type A_ξ is a control u_1 with $L = 1$ and $\xi = t_1^{(1)} \geq 0$; a control of type B is a control with two active portions ($L = 2$).

The set Ω of all p-trajectories defined for each concrete problem by the corresponding boundary conditions may be divided* into the following subsets:

a) Ω_1, consisting of p-trajectories on which $R(p^0, t)$ is a monotonic function of the variable t $(0 \leq t \leq T)$ in which $R(p_1^0, t) > 0$;

b) Ω_2, consisting of trajectories on which $t_0 \in [0, T]$ $(2t_0 \leq T - t_a)$;

c) Ω_3, which is defined by the conditions: $t_0 \in [0, T]$ $(T + t_a > 2t_0 > T - t_a)$;

d) Ω_4, defined by the conditions: $t_0 \overline{\in} [0, T]$, $\dot{R}(p^0, t) < 0$ (Ω_4^0) and $t_0 \in [0, T]$, $2t_0 > T + t_a$ (Ω_4^1).

For problem 2 (sec. 4) such a subdivision is accomplished as in Fig. 6 for arbitrary initial point $(-T, p_2^0)$; the set of all possible p-trajectories issuing at the moment $t = 0$ from the points of the straight line $p_1 = -T$ $(-\infty < p_2^0 < \infty$, $-\infty < p_3^0 < \infty)$, fill out the semi-strip $T < p_1 \leq 0$.

An A_0-control may clearly arise which satisfies the necessary condition (5), on any p-trajectory belonging to the set Ω. As for the question of whether there exists types of controls different from A_0-controls on p-trajectories (for the problem with fixed finite mass), the following assertion provides the answer (this is a consequence of Theorem 1).

Theorem 2. If $U_{1p}^{(i)}$ $(i = 1, \ldots, 4)$ is the set of all optimal controls u_1 possible for an arbitrary p-trajectory, lying in the set Ω_i $(i = 1, \ldots, 4)$, and depending on the parameter p_3^0, then:

a) $U_p^{(1)}$ is composed of controls of type A_ξ (in particular, type A_0), $U_{1p}^{(2)}$ of controls of types B, A_ξ, and A_0, and $U_{1p}^{(3)}$ of controls of type B (in the limit a type A_0), and $U_{1p}^{(4)}$ exhausts the controls of type A_0;

b) the set $U_{1p}^{(i)}$ is covered when the parameter p_3^0 varies in a finite region $E_p^{(i)}$, which is uniquely determined for any p-trajectory.

If one considers Ω^* as the union of the set of p-trajectories Ω_2, Ω_3, and Ω_4^1, then Theorem 2 may be extended to the case of a problem in which the finite value of the mass is not fixed.

In Fig. 5a a control u_1 from some set $U_{1p}^{(1)}$, for which the region $E_p^{(1)}$ is determined by the condition $p_{3(1)}^0 \leq p_3^0 \leq p_{3(3)}^0$ is represented. In Fig. 5b a control u_1 from the set $U_{1p}^{(2)}$ is shown. Inside the region $E_p^{(2)}$, $p_{3(1)}^0 \leq p_3^0 \leq p_{3(4)}^0$ there exist critical points such that the type of control changes upon passing through them; for $p_{3(1)}^0 \leq p_3^0 < p_{3(5)}^0$ there occurs a regime of A_ξ-control; for $p_{3(5)}^0 < p_3^0 < p_{3(4)}^0$ a regime of B-control; as $p_3^0 \to p_{3(4)}^0$ the second active portion degenerates, and for $p_3^0 = p_{3(4)}^0$ there is established an A_0-control; for $p_3^0 < p_{3(1)}^0$ boundary condition (14) is not satisfied; when p_3^0 increases beyond $p_{3(4)}^0$, the type of regime undergoes no further changes.

From Theorem 2, in particular, it follows that the optimal control in problem 4 (Sec. 4) is an A_0-regime (since the set of all possible p-trajectories for this problem is of type Ω_4^0). The fact noted earlier without proof was used in the papers of S. Ross [11], B. Fried, and J. Richardson [13].

*For the case of problems with fixed finite mass.

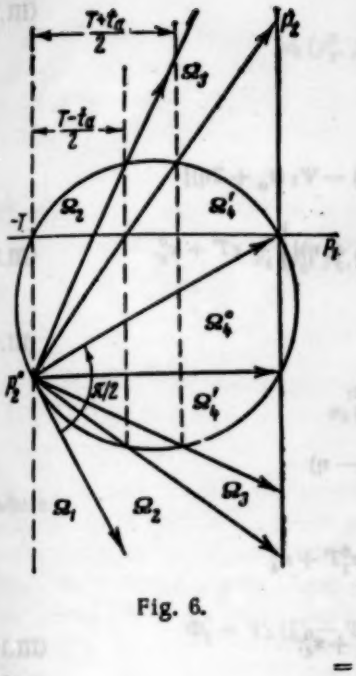


Fig. 6.

APPENDIX III

Integration of Equations (8) for Problem 2 (sec. 4)

We shall determine the location of the point $x(T) = \{x_1(T), \ldots, x_5(T)\}$ of the space X into which some optimal control $u(t) = \{u_1(p_2^0, p_3^0, \zeta), u_2(p_2^0, p_3^0)\}$ takes system (1) from the fixed initial position x^0 in a fixed time $t = T$; i.e., we shall find the integral of the system (8) of differential equations in a form depending on the parameters p_2^0, p_3^0, ζ. The values of this integral is used to determine the optimal control $u^*(t)$ (i.e., p_2^{0*}, p_3^{0*}, ζ^*) by the prescription of the initial and boundary conditions with the help of the devices described above (sec. 3, II).

A. We first examine the case of B-control, being the most general. For the parameter ζ, reckoned in this case from the point t_0, we introduce the special notation $\zeta \equiv \eta$. Dividing the interval of integration $[0, T]$ into the parts $[0, t_0 - \eta)$, $(t_0 - \eta, t_0 + \eta)$, $(t_0 + \eta, t_a + 2\eta)$, $(t_a + 2\eta, T]$, we integrate the functions $\dot{x}_i(t)$ separately in each part.

In the first active part $0 \le t < t_0 - \eta$

$$x_1(t) = -\alpha c \int_0^t \frac{p_1(\xi)\,d\xi}{(1-\alpha\xi)\sqrt{p_1^2(\xi)+p_2^2(\xi)}} + x_1^0$$

$$= \frac{c}{\sqrt{a}}\left\{\Psi_1(0) - \Psi_1(t) - \frac{B}{\sqrt{a_1}}[\Psi_2(t) - \Psi_2(0)]\right\} + x_1^0, \qquad \text{(III.1)}$$

where

$$\Psi_1(\xi) = -\operatorname{Ar\,sh}\frac{b+\xi}{\sqrt{\Delta}}, \qquad \Psi_2(\xi) = \operatorname{Ar\,sh}\frac{\sqrt{\frac{\alpha}{1-\alpha\xi}} - b_1}{\sqrt{\Delta_1}}, \qquad b = -\frac{p_2^0 p_3^0 + T}{a},$$

$$B = \frac{1}{\alpha} - T, \qquad a = p_3^{02} + 1, \qquad d = \frac{p_3^{02}}{a}, \qquad \Delta = D^2 d(1-d), \qquad \text{(III.2)}$$

$$\Delta_1 = \frac{\Delta}{a_1^2}, \qquad a_1 = \frac{1}{a}[B^2 + p_3^{02}(D-B)^2], \qquad b_1 = \frac{\beta + b}{a_1}, \qquad \beta = \frac{1}{\alpha}.$$

In the case $\Delta = \Delta_1 = 0$ (which we shall not pursue beyond this point), in place of (III.1) we have

$$x_1(t) = \frac{c}{\sqrt{a}}\left[\ln\frac{t+b}{b} - \frac{B}{\sqrt{a_1}}\ln\frac{\sqrt{\frac{\alpha}{1-\alpha t}} - b_1}{\alpha - b_1}\right] + x_1^0. \qquad \text{(III.3)}$$

On the first passive portion $u_1(t) = 0$ $(t_0 - \eta < t < t_0 + \eta)$ and

$$x_1(t) = x_1(t_0 - \eta) = \frac{c}{\sqrt{a}}\left\{\Psi_1(0) - \Psi_1(t_0 - \eta) - \frac{B}{\sqrt{a_1}}[\Psi_2(t_0 - \eta) - \Psi_2(0)]\right\} + x_1^0. \qquad \text{(III.4)}$$

On the second active portion, $t_0 + \eta < t < t_a + 2\eta$, the horizontal velocity obtains an increase

$$\Delta x_{1\eta} = -\alpha c \int_{t_0+\eta}^{t_a+2\eta} \frac{p_1(\xi)\,d\xi}{(1+2\alpha\eta-\alpha\xi)\sqrt{p_1^2(\xi)+p_2^2(\xi)}}$$

$$= \frac{c}{\sqrt{a}}\left\{\Psi_1(t_0+\eta) - \Psi_1(t_a+2\eta) - \frac{B_\eta}{\sqrt{a_{1\eta}}}[\Psi_2(t_a+2\eta) - \Psi_2(t_0+\eta)]\right\}, \qquad \text{(III.5)}$$

where

$$\Psi_2(\xi) = \operatorname{Ar\,sh}\frac{\sqrt{\frac{\alpha}{1-\alpha(\xi-2\eta)}} - b_{1\eta}}{\sqrt{\Delta_{1\eta}}}, \qquad B_\eta = B + 2\eta, \qquad \beta_\eta = \beta + 2\eta,$$

$$b_{1\eta} = \frac{\beta_\eta + b}{a_{1\eta}}, \qquad \Delta_{1\eta} = \frac{\Delta}{a_{1\eta}^2}, \qquad a_{1\eta} = \frac{1}{a}[B_\eta^2 + p_3^{02}(D - B_\eta)^2]. \qquad \text{(III.6)}$$

Finally

$$x_1(p_2^0,\ p_5^0,\ \eta,\ T) = x_1(t_0-\eta) + \Delta x_{1\eta}. \tag{III.7}$$

Analogously for $x_2(p_2^0, p_5^0, \eta, T)$ we obtain

$$x_2(p_2^0, p_5^0, \eta, T) = \frac{c}{\sqrt{a}}\left\{-p_5^0[\Psi_1(0) - \Psi_1(t_0-\eta) + \Psi_1(t_0+\eta) - \Psi_1(t_a+2\eta)]\right.$$

$$\left. - \frac{C - p_5^0 B}{\sqrt{a_1}}[\Psi_2(t_0-\eta) - \Psi_2(0)] - \frac{C - p_5^0 B_\eta}{\sqrt{a_{1\eta}}}[\Psi_2(t_a+2\eta) - \Psi_2(t_0+\eta)]\right\} - gT + x_2^0, \tag{III.8}$$

where

$$C = p_2^0 - p_5^0 T. \tag{III.9}$$

The values of the range and height are determined in the following manner:

$$x_4(p_2^0, p_5^0, \eta, T) = \int_0^{t_0-\eta} x_1(\xi)\,d\xi + (T - t_0 + \eta)\,x_1(t_0-\eta)$$

$$+ \int_{t_0+\eta}^{t_a+2\eta} \Delta x_{1\eta}(\xi)\,d\xi + (T - t_a - 2\eta)\,\Delta x_{1\eta}(t_a+2\eta) + x_1^0 T + x_4^0$$

$$= \frac{c}{\sqrt{a}}[\Phi_1 + B(\Phi_2 + \Phi_4) + B_\eta(\Phi_3 + \Phi_5)] + x_1^0 T + x_4^0; \tag{III.10}$$

$$x_5(p_2^0, p_5^0, \eta, T) = \frac{c}{\sqrt{a}}[-p_5^0\Phi_1 + (C - p_5^0 B)(\Phi_2 + \Phi_4) + (C - p_5^0 B_\eta)(\Phi_3 + \Phi_5)]$$

$$- \frac{gT^2}{2} - x_2^0 T + x_5^0, \tag{III.11}$$

where

$$\Phi_1 = (b+T)[\Psi_1(0) - \Psi_1(t_0-\eta) + \Psi_1(t_0+\eta) - \Psi_1(t_a+2\eta)]$$

$$+ \sqrt{b^2+\Delta} - \sqrt{(b+t_0-\eta)^2+\Delta} + \sqrt{(b+t_0+\eta)^2+\Delta} - \sqrt{(b+t_a+2\eta)^2+\Delta},$$

$$\Phi_2 = \frac{B}{\sqrt{a_1}}[\Psi_2(t_0-\eta) - \Psi_2(0)], \qquad \Phi_3 = \frac{B_\eta}{\sqrt{a_{1\eta}}}[\Psi_2(t_a+2\eta) - \Psi_2(t_0+\eta)],$$

$$\Phi_4 = \frac{1}{\sqrt{a_1 a_2}}\left[\operatorname{Ar\,sh}\frac{\frac{1-\alpha(t_0-\eta)}{\alpha} - b_2}{\sqrt{\Delta_2}} - \operatorname{Ar\,sh}\frac{\frac{1}{\alpha} - b_2}{\sqrt{\Delta_2}}\right], \tag{III.12}$$

$$\Phi_5 = \frac{1}{\sqrt{a_{1\eta} a_{2\eta}}}\left[\operatorname{Ar\,sh}\frac{\frac{1-\alpha t_a}{\alpha} - b_{2\eta}}{\sqrt{\Delta_{2\eta}}} - \operatorname{Ar\,sh}\frac{\frac{1-\alpha(t_0-\eta)}{\alpha} - b_{2\eta}}{\sqrt{\Delta_{2\eta}}}\right],$$

$$a_2 = b_1^2 + \Delta_1,$$

$$b_2 = \frac{b_1}{a_2}, \quad \Delta_2 = \frac{\Delta_1}{a_2^2}, \quad a_{2\eta} = b_{1\eta}^2 + \Delta_{1\eta}, \quad b_{2\eta} = \frac{b_{1\eta}}{a_{2\eta}}, \quad \Delta_{2\eta} = \frac{\Delta_{1\eta}}{a_{2\eta}^2}.$$

B. For optimal controls of type A the entire interval [0, T] is divided into three portions: $0 \le t < \xi$, where $u_1(t) = 0$; $\xi < t < t_a + \xi$, where $u_1(t) = 1$, and $t_a + \xi < t \le T$, on which $u_1(t) = 0$. For the parameter ζ in this case we take $\zeta \equiv \xi$, the beginning time for the active portion. The following relations for the A-regime are consequences of formulas (III.7), (III.8), (III.9), (III.10), and (III.11):

$$x_1(p_2^0, p_5^0, \xi, T) = \frac{c}{\sqrt{a}}\left(\Phi_1^* - \frac{B_\xi}{\sqrt{a_{1\xi}}}\Phi_2^*\right) + x_1^0, \tag{III.3}$$

$$x_2(p_2^0, p_5^0, \xi, T) = \frac{c}{\sqrt{a}}\left(-p_5^0\Phi_1^* - \frac{C - p_5^0 B_\xi}{\sqrt{a_{1\xi}}}\,\Phi_2^*\right) - gT + x_2^0, \tag{III.14}$$

$$x_4(p_2^0, p_5^0, \xi, T) = \frac{c}{\sqrt{a}}\left[b + T\Phi_1^* + \frac{B_\xi^2}{\sqrt{a_{1\xi}}}\,\Phi_2^* + \Phi_3^* + \frac{B_\xi\Phi_4^*}{\sqrt{a_{1\xi}a_{2\xi}}}\right] + x_1^0 T + x_4^0, \tag{III.15}$$

$$x_5(p_2^0, p_5^0, \xi, T) = \frac{c}{\sqrt{a}}\left[-p_5^0 b + T\Phi_1^* + \frac{(C - p_5^0 B_\xi)\,B_\xi\Phi_2^*}{\sqrt{a_{1\xi}}}\right.$$

where

$$\left. - p_5^0\Phi_3^* + \frac{(C - p_5^0 B_\xi)\,\Phi_4^*}{\sqrt{a_{1\xi}a_{2\xi}}}\right] - \frac{gT^2}{2} + x_2^0 T + x_5^0, \tag{III.16}$$

$$\Phi_1^* = \Psi_1(\xi) - \Psi_1(t_a + \xi), \qquad \Phi_2^* = \operatorname{Ar\,sh}\frac{\dfrac{\alpha}{1 - \alpha t_a} - b_{1\xi}}{\sqrt{\Delta_{1\xi}}} - \operatorname{Ar\,sh}\frac{\alpha - b_{1\xi}}{\sqrt{\Delta_{1\xi}}},$$

$$\Phi_3^* = \sqrt{(b + \xi)^2 + \Delta} - \sqrt{(b + \xi + t_a)^2 + \Delta},$$

$$\Phi_4^* = \operatorname{Ar\,sh}\frac{\dfrac{1 - \alpha t_a}{\alpha} - b_{2\xi}}{\sqrt{\Delta_{2\xi}}} - \operatorname{Ar\,sh}\frac{\dfrac{1}{\alpha} - b_{2\xi}}{\sqrt{\Delta_{2\xi}}}. \tag{III.17}$$

In (III.13- 17) the quantities with index ξ are determined by the formula in sec. A, Appendix III, in which one should always replace η by ξ, in which $B_\xi = B + \xi$, $\beta_\xi = \beta + \xi$.

LITERATURE CITED

1. Boltyanskii, V. G., Gamkrelidze, R. V., and Pontryagin, L. S., "Theory of optimal processes. I", Izv. AN SSSR, ser. matem., Vol. 24, No. 1, 1960.
2. Rozonoér, L. I., "Pontryagin's maximum principle in the theory of optimal systems. II", Avtomatika i Telemekhanika, Vol. XX, No. 11, 1959.
3. Rozonoér, L. I., "L. S. Pontryagin's maximum principle in the theory of optimal systems. I", Avtomatika i Telemekhanika, Vol. XX, No. 10, 1959.
4. Gorelov, Yu. A., "On two classes of plane extremal motions of rockets in vacuum", Prikl. Matem. i Mekh., Vol. XXIV, No. 2, 1960.
5. Miele A., Capellari J. Topics in Dynamic Programming for Rockets. Zeitschrift für Flugwissenschaften, Bd. I, Nr. 1, 1959.
6. Okhotsimskii, D. E., Eneev, T. M., "Certain variational problems connected with orbiting an artificial Earth satellite", Usp. fiz. nauk, Vol. LXIII, No. 1a, 1957.
7. Miele A. An Extension of the Theory of the Optimum Burning Program for the Level Flight of a Rocket Powered Aircraft. Journ. Aeronaut. Sci., Vol. 24, No. 12, 1957.
8. Leitmann G. On a Class of Variational Problems in Rocket Flight. Journ. Aero/Space Sci. Vol. 26, No. 9, 1959.
9. Lowden D. Dynamic Problems of Interplanetary Flight. Aeronautical Quarterly, Vol. 6, p. 3, 1955.
10. Lowden D. Optimal Rocket Trajectories. Jet Propulsion, Vol. 27, No. 12, 1957.
11. Ross S. Composite Trajectories Yielding Maximum Coasting Apogee Velocity. ARS Journal, Vol. 26, No. 11, 1959.
12. Fried B. D. On the Powered Flight Trajectory of an Earth Satellite. Jet. Propulsion, Vol. 27, No. 6, 1957.
13. Fried B. D., Richardson J. Optimal Rocket Trajectories. Journ. Appl. Phys., Vol. 27, No. 8, 1956.

AUTOMATIC OPTIMIZATION OF SPACE DISTRIBUTION. III

L. N. Fitsner (Moscow)

Translated from Avtomatika, Vol. 22, No. 8,
pp. 1002-1012, August 1961
Original article submitted December 12, 1960

Design principles for the elements of equipment intended for the automatic synthesis of space distribution optimization systems considered in [1] are presented.

1. On the Automatic Synthesis of Automatic Search Systems

It was noted in [1] that in the solution of certain problems connected with the design of space distribution systems, it is not always convenient to use analytic methods. In particular, these problems are: a) the finding of the optimal form of test action, b) the synthesis of circuits operating in accordance with the chosen form of action, c) the determination of the coordinate converter coefficients, d) the choice of optimal time intervals between measurements of increments for a given range of variation of the inertial element parameters and the nonlinear characteristics of the system, e) the determination of the optimal step magnitude in the presence of random noise, and others.

For the solution of given problems a set of equipment has been developed which we have called the automatic synthesizer.

The question of the necessity for developing apparatus for the automatic synthesis of automatic control systems was, as far as we know, first posed by A. A. Fel'dbaum [2]. Together with R. I. Stakhovskii he developed one of the basic units of the synthesizer – a multichannel automatic optimizer which permits determination of the extreme values of a function of 12 variables in the presence of supplementary constraints [3].

The automatic synthesizer contains the following basic units:

a) multichannel automatic optimizer for finding the extreme values of a function (in the presence of one or several extremals for the function);

b) a linear filter, whose parameters can be varfied by the action of automatic optimizer voltages;

c) a controlled nonlinear converter with a single input;

d) a controlled nonlinear converter with two or more inputs;

e) a controlled model of the object, including both linear and nonlinear elements;

f) a voltage memory unit [4, 5].

The design principles for the above units are presented below.

The automatic synthesizer, generally speaking, permits determination of the best structure for the control apparatus of a given class of automatic systems. However, one of the most important fields of application of the synthesizer is the solution of problems connected with the design of optimal automatic search systems and, in particular, systems for the automatic optimization of a space distribution. The synthesis of these systems is connected with the solution of a number of complex variational problems for which the methodology of solution is relatively weakly developed at the present time.

2. Automatic Optimizer

The automatic optimizers proposed by A. A. Fel'dbaum and R. I. Stakhovskii permit carrying out the search for extremal values of a function F of several (up to 10-12) variables ($x_1, x_2, \ldots, x_n$) by the method of steepest descent or the gradient method [2, 6]. Our investigations have shown [6] that a given form of optimizer will permit, as a rule, a fairly rapid and accurate determination of the extremal values of a function even in the presence of

several supplementary constraints on the region of permissible variation of $x_1, x_2, \ldots, x_n$.

The present optimizer can be used when the function F has a single extremum. In the presence of several extreme values of the function (or in the absence of a proof on the presence of a single extremum) it becomes necessary to examine the entire range of variation of the variables. An optimizer has been worked out for such investigations, operating on a scanning method.

It should be remarked that in scanning, the search time T_s strongly increases with the number of variables n, and therefore this method can be used only for small n. Thus, if each of the n variables takes on m values during the scanning process and the frequency of change of variable in the model of the object is equal to f, we have

$$T_s = m^n / f. \qquad (1)$$

Let m = 10, f = 1000 cps, and n = 6; then T_s = 16.7 min.

A block diagram of the optimizer AO is given in Fig. 1. The optimizer includes the drive unit 1, a comparator 2, a memory unit 3, a limiting block 4, and a clock generator 5. The object of search O (or model of the object) is shown at the right.

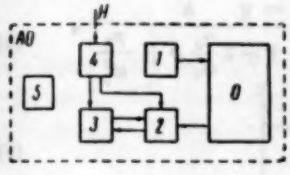


Fig. 1.

The values of the input variables $x_1, x_2, \ldots, x_n$ are applied to the object inputs from the n drive units, equal in number to the object inputs. The drive unit is an operational amplifier [7], whose input voltage is taken from the cathodes of a series chain of cold-cathode thyratrons. The number of thyratrons in the chain is defined by the number m of values taken on by the quantity x_i during the scanning process.

The thyratrons in the chain are fired in turn by the clock pulse generators; this varies the voltage at the output of the operational amplifier.

In the optimizer model developed by us each of the variables can take on 10 values. Consequently, the ratio of frequencies of successive variables is 10. The maximum frequency is taken equal to 1000 cps.

The comparator obtains voltages from the object O and the memory unit.

In searching for a maximum of the function F, an instruction for storing a new value is given only if the voltage at the output of O exceeds the voltage in the memory unit. Then an instruction is given for storing both the new value of the output function of the model F, and the new values $x_1, x_2, \ldots, x_n$, corresponding to the maximum of the function F in the region investigated. The memory unit permits storing voltages to two significant figures. It consists of an electronic servomechanism using cold-cathode thyratrons.

Sometimes, from the conditions of the problem, the search for an extremum of the function F may be carried out only in a region for which the function H $(x_1, x_2, \ldots, x_n)$ does not exceed given limits.

In the automatic optimizer circuit a limiting unit is included which permits storing of new extremum coordinates only when the given conditions are observed.

3. Controlled Linear Filter (CLF)

The controlled linear filter consists of linear networks whose parameters can be varied by the action of external potentials obtained from the automatic optimizer. The need for such devices arises in the automatic synthesis of automatic search systems which satisfy the given requirements with regard to search dynamics and noise effects in the best manner.

The CLF contains amplifiers with variable gain factors, delay networks with variable time constants, and tuned networks, permitting variation of the natural frequency and decrement. These networks of the CLF can be interconnected in any arbitrary manner.

For the variable gain amplifier we use a multiplying circuit MC (Fig. 2a). Here the gain factor

$$k = V_{out}/V_{in} = \alpha V_c \qquad (2)$$

is proportional to the control voltage V_c (α is the proportionality factor of MC).

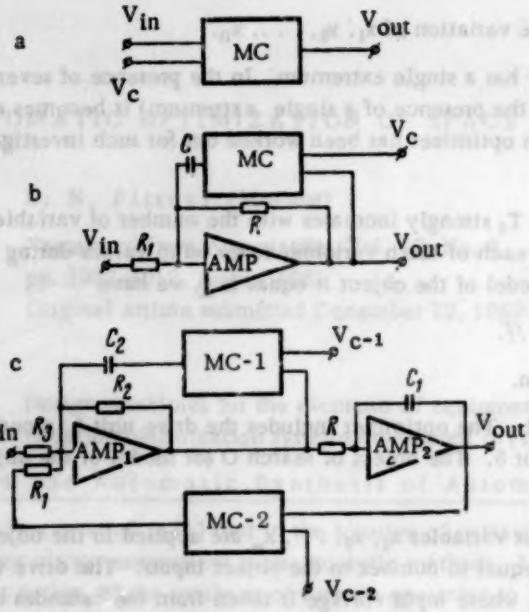


Fig. 2.

A block diagram of the inertial circuit is given in Fig. 2b. Here AMP is an operational amplifier. The transfer function of such a circuit is

$$K(p) = \frac{V_{out}(p)}{V_{in}(p)} = -\frac{k}{1+pT}. \quad (3)$$

The gain factor of the circuit $k = R/R_1 = \text{const}$; the time constant

$$T = RC\alpha V_c \quad (4)$$

can be varied under the action of the voltage V_c.

The circuit of a tuned network (Fig. 2c) contains two operational amplifiers AMP_1 and AMP_2 and two multiplier circuits MC-1 and MC-2. Its transfer function is

$$K(p) = \frac{V_{out}(p)}{V_{in}(p)} = \frac{k}{1+2\frac{\alpha_0}{\omega_0}p+\frac{p^2}{\omega_0^2}}, \quad (5)$$

where

$$k = \frac{R_1}{R_3\alpha V_{c-2}}, \quad (6)$$

$$2\frac{\alpha_0}{\omega_0} = \frac{R_1RC_1}{R_2\alpha V_{c-2}}, \quad (7)$$

$$\omega_0 = \sqrt{\frac{V_{c-2}}{V_{c-1}R_1C_2RC_1}}. \quad (8)$$

The parameters of this circuit (k, α_0, ω_0) can be varied under the action of voltages V_{c-1} and V_{c-2}.

To construct the controlled linear networks (CLF) a simple MC [8] was used, permitting variation of the linear circuit parameters over wide limits.

4. Controlled Nonlinear Converter with Single Input (CNC-1)

It is well-known [9] that for the realization of an optimal search law the extremal system should include nonlinear converters with one or several inputs, which are used for reproducing nonlinear functional dependencies among the variables. The forms of these dependencies can be found analytically only in the simplest systems [9]. In more complicated cases the solution of such problems must be carried out using a synthesizer.

Let us denote the input value of a nonlinear converter by x and its output quantity by y. The form of the function

$$y = f(x) \quad (9)$$

is not known in advance and, as noted above, must be found by the synthesizer in the process of automatic search.

In the problem of the controlled nonlinear converter with single input (CNC-1) there is included change of the form f under the influence of the automatic optimizer output voltages ($y_1, y_2, \ldots, y_n$). There are 11 such voltages in the CNC-1 which we have developed. By their means the function is reproduced at 11 points ($x_1, x_2, \ldots, x_{11}$) (Fig. 3).

The function is approximated by linear segments at equal intervals between these points.

A block diagram of the CNC-1 is given in Fig. 4. The converter can be divided into two subunits; the subunit of stepwise approximation (to the left of the broken line) and the linear approximator. The input voltage to the converter x is applied through the voltage dividers to the resistances R at the inputs to the trigger circuits T_1-T_{11}. The voltage dividers are calculated so that each of the trigger circuits will operate at determined values of x (the trigger circuit T_1 with $x \geq x_1$, T_2 with $x \geq x_2$, etc.).

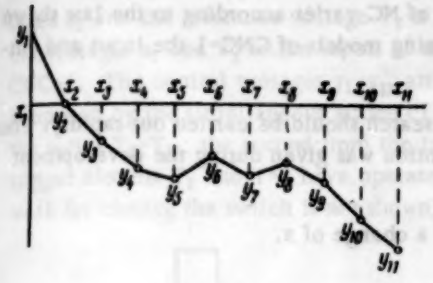


Fig. 3.

The trigger circuits control the electronic switches K_1-K_{11}. In the circuit full lines indicate switches opened by the trigger circuits, broken lines, switches closed by the trigger circuits. Each of the trigger circuits (except the end ones) operates two switches simultaneously. With this circuit, one of the voltages y_1-y_{11} with odd index (y_o) is applied to the input of amplifier AMP_1 and with even index (y_e) to the input of AMP_2.

The values of the converter output voltages (y_{2n-1}, y_{2n}, y_{2n+1}) for three values of x (x_{2n-1}, x_{2n}, x_{2n+1}) are given in Fig. 5. In the intervals between points the function is approximated by line segments

$$y = y_{2n} + (y_{2n-1} - y_{2n}) \left| \frac{x_{2n} - x}{x_{2n} - x_{2n-1}} \right| \quad \text{for} \quad x_{2n-1} < x < x_{2n} \tag{10}$$

and

$$y = y_{2n} + (y_{2n+1} - y_{2n}) \left| \frac{x_{2n} - x}{x_{2n} - x_{2n+1}} \right| \quad \text{for} \quad x_{2n} < x < x_{2n+1}. \tag{11}$$

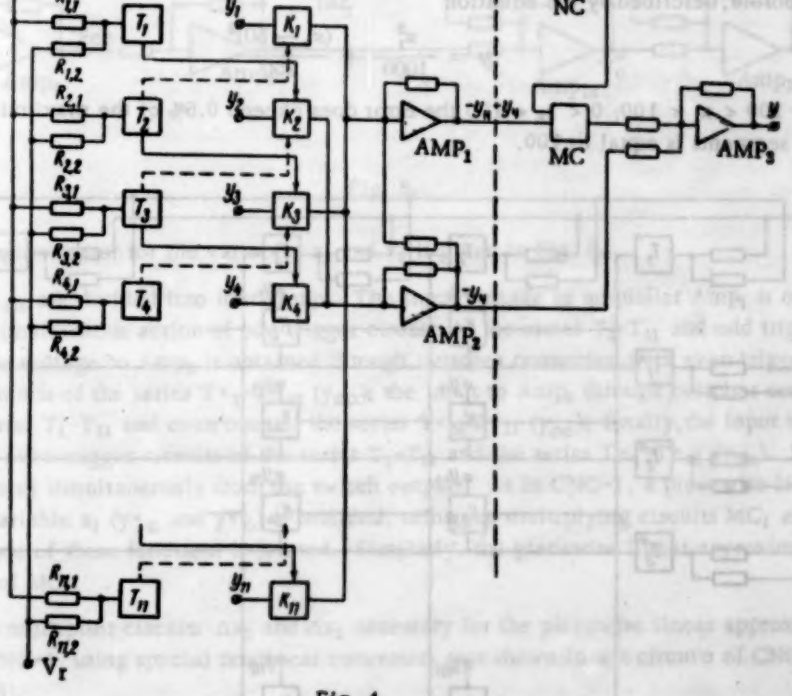


Fig. 4.

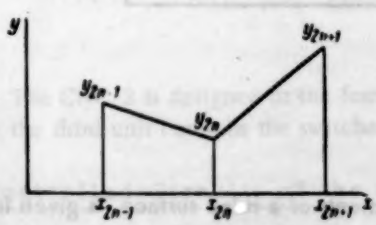


Fig. 5.

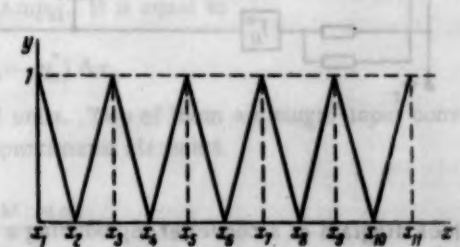


Fig. 6.

According to (10) and (11) the approximator circuit (Fig. 4) includes the multiplier circuit MC, the special nonlinear converter NC, and the output amplifier AMP_3. The output voltage of NC varies according to the law shown in Fig. 6. Here $\Delta = |x_{2n-1} - x_{2n}| = |x_{2n} - x_{2n+1}| =$ const. In functioning models of CNC-1 the input and output voltage scales cover the range ± 100 v.

For the purpose of shortening the duration of process optimization the search should be carried out rapidly. Therefore, the CNC-1 should have a wide frequency band. In this connection attention was given during the development work to shortening the transients in the trigger circuits and switches.

In the model developed the switches operated within 8-10 μ sec after a change of x.

5. Controlled Nonlinear Converter with Two Inputs (CNC-2)

The output voltage of the nonlinear converter with two inputs (CNC-2) is a nonlinear function of input voltages x_1 and x_2

$$y = f(x_1, x_2). \tag{12}$$

In CNC-2 as in CNC-1 the form of the function f is found during the solution of the problem in the automatic optimization system.

The CNC-2 developed reproduces a function of two variables by its approximation by segments of ruled surfaces. Before developing the converter it was necessary to estimate the error caused by the approximating function, for the purpose of finding an acceptable number of approximating segments. It is shown in the Appendix that in approximating an elliptical paraboloid, described by the equation

$$y = \frac{x_1^2}{1000} + \frac{(x_2 - 50)^2}{250}, \tag{13}$$

on the segment $-100 < x_1 < 100$; $0 < x_2 < 100$ the error does exceed 0.5% of the maximum value of y, if the number of approximating segments is equal to 100.

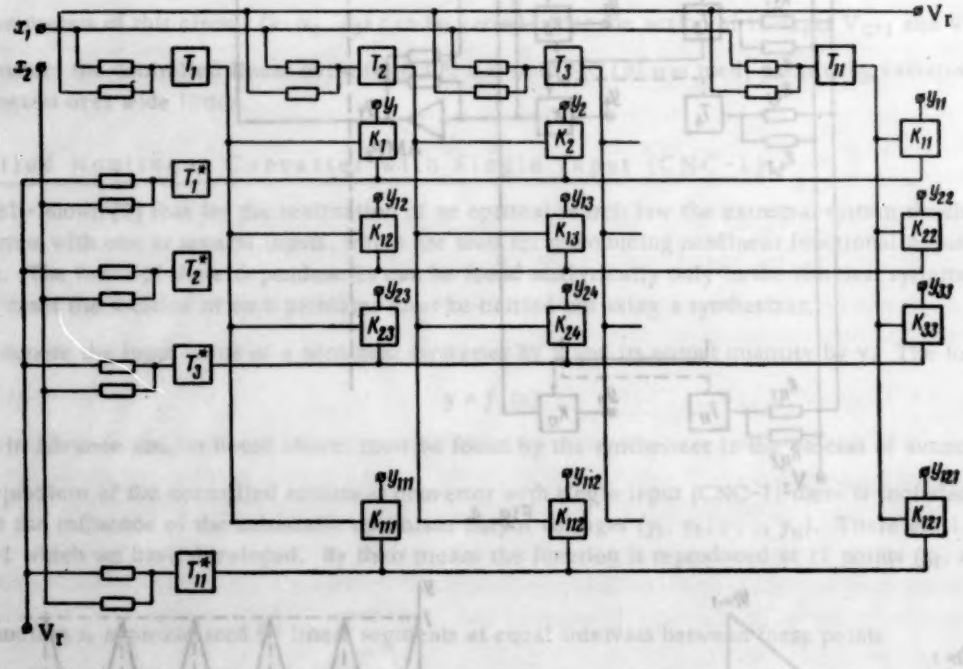


Fig. 7

The block diagram of a converter reproducing a function by 100 segments of a ruled surface, is given in Fig. 7 (subunit of stepwise approximation) and in Fig. 8 (liner approximator).

The stepwise approximation subunit consists of two series of trigger circuits (T_1-T_{11} and T^*_1-T^*_{11}). The voltages x_1 and the reference voltage V_r is applied to the inputs of trigger circuits T_1-T_{11} through voltage dividers, and the voltages x_2 and V_r to the inputs of trigger circuits T^*_1-T^*_{11}. These trigger chains operate in the same way as in CNC-1. The control voltages y_1-y_{121} are applied to the inputs of the electronic switches K_1-K_{121}. Each of the 121 switches is opened by the simultaneous action of two voltages. One of these is applied to the switch from a trigger circuit of the series T_1-T_{11}, the second from the trigger circuit of series T^*_1-T^*_{11}. Thus, for example, switch K_{24} opens only if trigger circuits T_2 and T^*_3 have operated. The switches close in the same way as in circuits CNC-1 (in Fig. 7 the network for closing the switch is not shown).

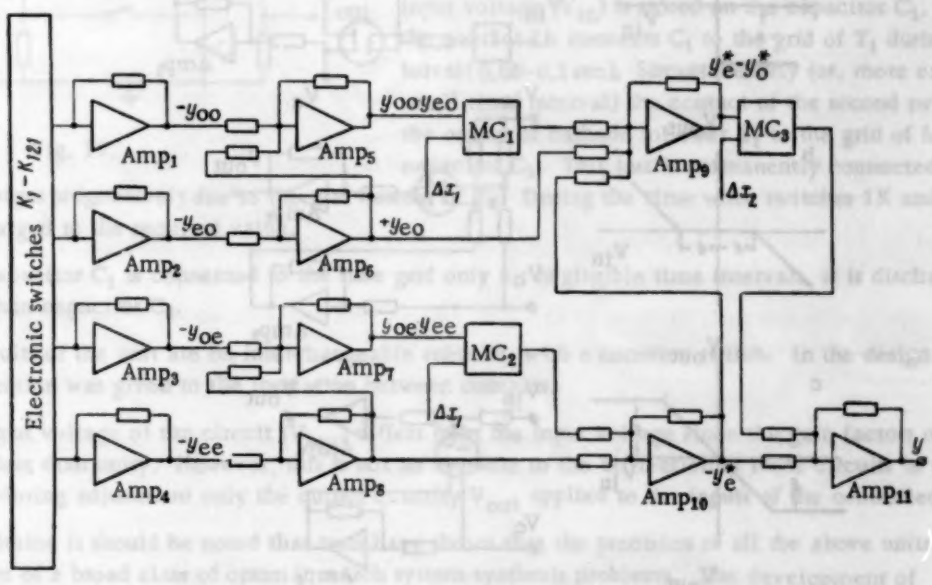


Fig. 8.

A circuit of the approximator for the variables x_1 and x_2 is given in Fig. 8.

The switches K_1-K_{121} are divided into four groups. The input voltage to amplifier Amp_1 is obtained through switches opened by the simultaneous action of odd trigger circuits in the series T_1-T_{11} and odd trigger circuits of the series T^*_1-T^*_{11} (y_{oo}); the voltage to Amp_2 is obtained through switches connected with even trigger circuits of the series T_1-T_{11} and odd circutis of the series T^*_1-T^*_{11} (y_{eo}); the input to Amp_3 through switches connected with odd trigger circuits of the series T_1-T_{11} and even ones of the series T^*_1-T^*_{11} (y_{oe}); finally, the input to Amp_4 through switches connected with even trigger circuits of the series T_1-T_{11} and the series T^*_1-T^*_{11} (y_{ee}). Four voltages are applied to the approximator simultaneously from the switch outputs. As in CNC-1, a piecewise linear approximation to two fucntions of the variable x_1 (y^*_e and y^*_o) is realized, using the multiplying circuits MC_1 and MC_2. At the output of Amp_9 the difference of these functions is formed. Similarly, the piecewise linear approximation for the variable x_2 is realized by means of MC_3.

The factors of the multiplier circuits Δx_1 and Δx_2 necessary for the piecewise linear approximation are formed in the same way as in CNC-1, using special nonlinear converters (not shown in the circuits of CNC-2), producing sawtooth voltages (cf. Fig. 6).

The output voltage of CNC-2 is taken from amplifier Amp_{11}. It is equal to

$$y = y^*_e + (y^*_o - y^*_e)\,\Delta x_2. \qquad (14)$$

The CNC-2 is designed in the form of three individual units. Two of them are single-input converters CNC-1, while the third unit contains the switches K_1-K_{121} and the approximator elements.

6. Controlled Circuits of the Search Object Model

In the design of the automatic search system considered in [1], it is important not only to synthesize the optimal control apparatus for the given object, but to find as well the admissible magnitudes of the object parameters for which

the search process would satisfy the prescribed requirements. For this it is necessary to be able to vary the parameters of the search object model under the action of automatic optimizer voltages.

The design of linear circuits with controllable parameters has been considered above.

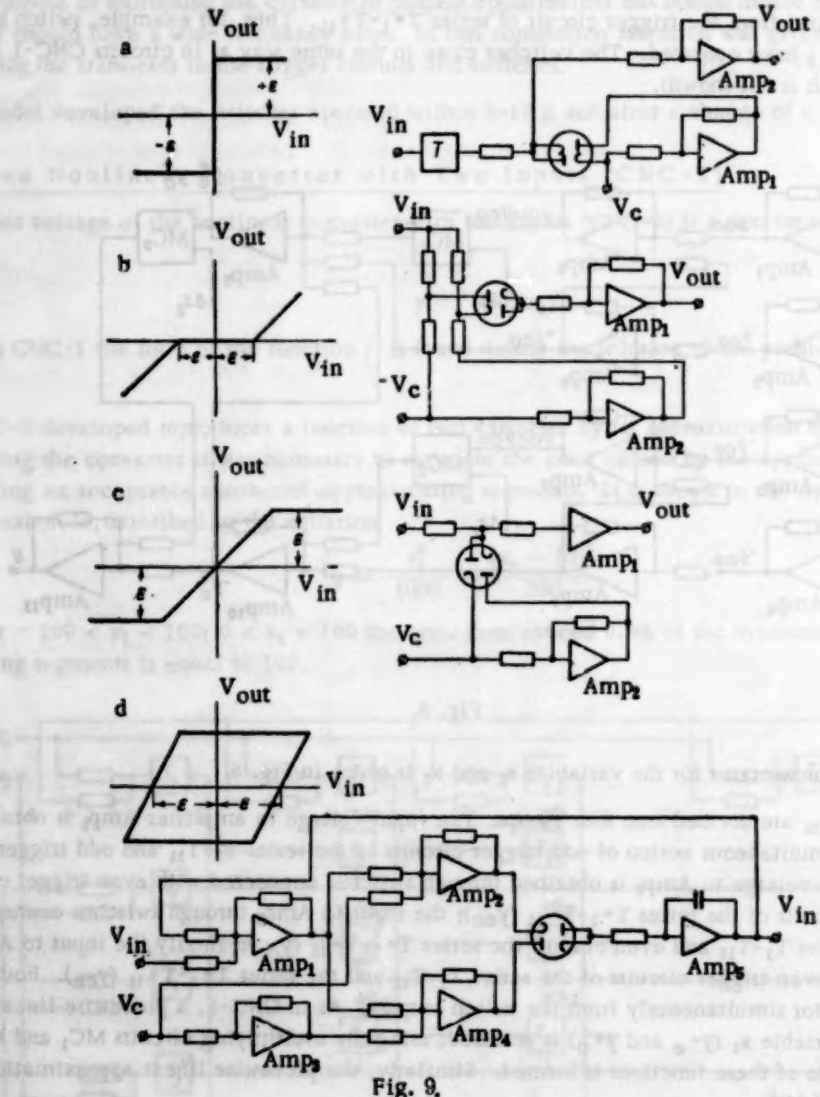


Fig. 9.

We shall only mention certain basic types of nonlinear dependencies occuring in the search object, and the circuits for producing them (Fig. 9). In the figure we have put: V_{in}-input voltage, V_{out}-output voltage, V_c-control voltage, ϵ - a variable parameter, Amp-operational amplifier, and T-trigger circuit. These circuits permit the reproduction of dry friction (a), deadzone (b), limiting (c),and backlash (d). The parameters ϵ in these circuits can be varied under the influence of the control voltages V_C obtained from the automatic optimizer.

If more complicated nonlinear dependencies occur in the object, it is possible to use the universal converter CNC-1 for their reproduction.

7. Voltage Memory Unit (MU-144)

Fairly often the number of quantities taking part in the optimization process exceeds the number of channels in the automatic optimizer. Thus, for example, to obtain acceptable precision in the reproduction of a nonlinear function of two variables the function is assigned at 121 points in the controlled two-input converter. It is also necessary to apply a substantial number of controlled voltages to other units of the automatic synthesizers as well. Automatic search in these cases can be carried out successively in separate groups so that the number of variables in each group does not exceed the number of optimizer channels.

With this form of search it is necessary to store the values of a large number of voltages and to apply them to the inputs of the controlled units. A sepcial unit MU-144 has been developed for storing voltages, consisting of 144 identical circuits (Fig. 10), each of which can store a single voltage in the range ± 100 v.

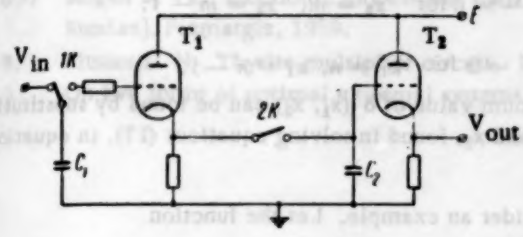


Fig. 10.

The circuit of the unit includes two cathode followers T_1 and T_2, polystyrene(C_1)and paper (C_2) capacitors, and switches 1K and 2K.

With the position of the switch shown in the diagram, the input voltage (V_{in}) is stored on the capacitor C_1. During operation the contact 1K connects C_1 to the grid of T_1 during a short time interval(0.05-0.1 sec). Simultaneously (or, more exactly, after a small time interval) the contact of the second switch 2K connects the output of cathode follower T_1 to the grid of follower T_2 and capacitor C_2. This last is permanently connected to the grid of T_2, and discharges progressively due to the grid current of T_2. During the time when switches 1K and 2K are closed it is again charged to the required value.

Since capacitor C_1 is connected to the tube grid only for negligible time intervals, it is discharged many times more slowly than capacitor C_2.

All circuits of the unit are on interchangeable subunits, with a common switch. In the design of the latter particular attention was given to the insulation between contacts.

The output voltage of the circuit (V_{out}) differs from the input voltage since the gain factors of the cathode followers are less than unity. However, this is not an obstacle to the utilization of these circuits in the optimizer system, since during adjustment only the output quantity V_{out} applied to the inputs of the controlled units is effective.

In conclusion it should be noted that tests have shown that the precision of all the above units is acceptable for the solution of a broad class of optimal search system synthesis problems. The development of the complex of equipment was carried out in the Institute of Automation and Remote Control of the Academy of Sciences of the U. S. S. R. The units CLF, CNC-1, CNC-2 and MU-144 were developed with the participation of K. B. Norkin. A. V. Torgashin and V. A. Yakovlev took part in the development of the automatic scanning optimizer.

APPENDIX

The error of approximating a function of two variables by segments of ruled surfaces

Above it was noted that with reproduction of a function $y = f(x_1, x_2)$ using CNC-2, the function is approximated by segments of ruled surfaces. One such segment is shown in Fig. 11. Here the points $M_1(x_{1n-1}, x_{2m}, y_1)$, $M_2(x_{1n}, x_{2m}, y_2)$, $M_3(x_{1n}, x_{2m-1}, y_3)$ and $M_4(x_{1n-1}, x_{2m-1}, y_4)$ lie simultaneously on the surface $f(x_1, x_2)$ and on the ruled surface.

The equation of the ruled surface is written in the form

$$ax_1x_2 + bx_1 + cx_2 + dy + l = 0, \tag{15}$$

where

$$\begin{aligned}
a &= y_2 - y_1 + y_4 - y_3,\\
b &= x_{2m}(y_3 - y_4) + x_{2m-1}(y_2 - y_1) + 2x_{2m}(y_1 - y_2),\\
c &= -x_{1n-1}(y_3 + y_2) + x_{1n}(2y_3 - y_4) + x_{1n}y_1,\\
d &= (y_3 - y_1)(x_{1n-1} - x_{1n}),\\
l &= x_{2m}y_4x_{1n} - x_{2m}x_{1n-1}y_3 - x_{1n-1}x_{2m-1}y_2 -\\
&\quad - y_1x_{2m-1}x_{1n} + x_{2m}x_{1n-1}y_2 - y_1x_{1n}x_{2m} + x_{2m}y_2 - y_1x_{1n}.
\end{aligned}$$

Let us write the equation of approximation error

$$\delta(x_1, x_2) = f(x_1, x_2) - y. \tag{16}$$

Here y are the heights of the points on the ruled surface.

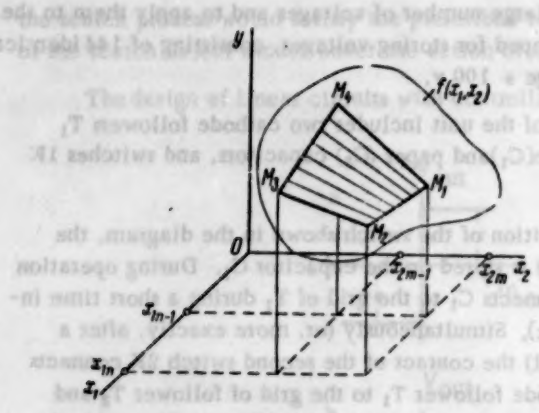


Fig. 11.

The values of x_1 and x_2 corresponding to the maximum error are found from the equation

$$\frac{\partial\delta(x_1, x_2)}{\partial x_1} = 0, \quad \frac{\partial\delta(x_1, x_2)}{\partial x_2} = 0$$

and $\dfrac{\partial\delta(x_1, x_2)}{\partial x_1} = 0$ for $x_2 = m, \quad x_2 = m - 1;$ (17)

$$\frac{\partial\delta(x_1, x_2)}{\partial x_2} = 0 \text{ for } x_1 = n, \; x_1 = n - 1.$$

The maximum value of $\delta(x_1, x_2)$ can be found by substituting the values of x_1 and x_2, found in solving equations (17), in equation (16).

Let us consider an example. Let the function

$$f(x_1, x_2) = \frac{x_1^2}{1000} + \frac{(x_2 - 50)^2}{250} \tag{18}$$

be approximated by 100 segments of a ruled surface. In this case the vaues of $f(x_1, x_2)$ should be assigned at 121 points.

Let us assume that x_1 can vary from -100 to $+100$, and x_2 from 0 to 100, and that the approximation region of variation of x_1 and x_2 is divided into equal segments. Then

$$x_{1n} = x_{1n-1} + 20, \quad x_{2m} = x_{2m-1} + 10. \tag{19}$$

In this example the coefficient of x_1x_2 in equation (15) is equal to zero, and the function $f(x_1, x_2)$ is approximated by plane segments.

In effect,

$$y_1 = \frac{x_{1n}^2}{1000} + \frac{(x_{2m} - 50)^2}{250}, \quad y_2 = \frac{(x_{1n} + 20)^2}{1000} + \frac{(x_{2m} - 50)^2}{250},$$

$$y_3 = \frac{(x_{1n} + 20)^2}{1000} + \frac{(x_{2m} - 40)^2}{250}, \quad y_4 = \frac{x_{1n}^2}{1000} + \frac{(x_{2m} - 40)^2}{250}.$$

$$a = y_2 - y_1 + y_4 - y_3 = 0.$$

According to (15), (16), and (17),

$$\frac{\partial\delta(x_1, x_2)}{\partial x_1} = \frac{\partial f(x_1, x_2)}{\partial x_1} - \frac{\partial y}{\partial x_1} = \frac{x_1}{500} + \frac{b}{d} = 0, \tag{20}$$

$$\frac{\partial\delta(x_1, x_2)}{\partial x_2} = \frac{\partial f(x_1, x_2)}{\partial x_2} - \frac{\partial y}{\partial x_2} = \frac{x_2}{125} - \frac{2}{5} + \frac{c}{d} = 0. \tag{21}$$

Hence $x_1 = x_{1n} + 10$, $x_2 = x_{2m} - 5$. Substituting (21) in (16), we find $\delta(x_1, x_2) = -0.2$. The maximum of $f(x_1, x_2)$ with variation of x_1 and x_2 in the given range is

$$\text{Max } f(x_1, x_2) = 20.$$

From this we find the error relative to the maximum value of the function, $\delta = -1\%$.

This magnitude of error can be reduced by a factor of 2 if at the nodal points of CNC-2 in place of $f(x_1, x_2)$ we put the function $f(x_1, x_2) + \delta/2$.

LITERATURE CITED

1. Fitsner, L. N., Automatic optimization of space distribution I, II. Avtomatika i telemekhanika, Vol. XXII, No. 1 and 7, 1961.
2. Fel'dbaum, A. A., Basic directions of development of digital equipment for automatic systems. Proceedings of the meeting of the Academy of Sciences of the U. S. S. R. on scientific problems of automation. Academy of Sciences, 1957.
3. Fel'dbaum, A. A., Automatic optimizer. Avtomatika i telemekhanika, Vol. XIX, No. 8, 1958.
4. Fitsner, L. N., Device for storage of electrical potential. Author's certificate No. 116680 with priority, Sept. 30, 1958.

5. Norkin K. B., Storage device for analog computers. Proceedings of a Conference of Young Scientists of the Institute of Automation and Remote Control of the Academy of Sciences of the U. S. S. R., 1959.
6. Stakhovskii R. I., Fitsner L. N., Shubin A. S., Automatic optimizers and their application for the solution of variational problems and automatic synthesis. First on Automatic control (Butterworth's Vol. 3, p. 141), 1961.
7. Kogan B. Ya., Electronic models and their application in the investigation of automatic control systems [in Russian]. Fizmatgiz, 1959.
8. Fitsner, L. N., Thyrite multiplier circuit. Priborostroenia, N. 4, 1956.
9. On two forms of optimal extremal systems. Avtomatika i telemekhanika, Vol. XXI. No. 8. 1960.

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

THE DETERMINATION OF OPTIMAL CHARACTERISTICS FOR AN EXTREMAL SYSTEM WITH RANDOM DISTURBANCES

N. V. Grishko (Moscow)

Translated from Avtomatika i Telemekhanika Vol. 22, No. 8,
pp. 1013-1026, August, 1961
Original article submitted July 26, 1960

An analytical expression is obtained for the mean square error in an extremal servosystem. Results of interest in engineering are derived.

1. Introduction

This article is devoted to the investigation of the extremal system described in [1]. Since certain extremal step systems of other types can be considered as particular cases of this system, the investigation has a general value. An analytical expression for the quality criterion of the system obtained by the author is used as a basis for the investigation. This expression is obtained in the first part of the article. In the second part, this expression will be used in the derivation of optimal characteristics for the extremal system considered.

The block diagram of the extremal system is shown in the diagram.

An oscillation x_2 is generated in the system, which is led to the input of the object and to one of the inputs of a multiplier. In the multiplier, the oscillation x_2 is first transformed into an oscillation x_{2K} and then multiplied with the output coordinate y of the object. In A, an averaging of this product is carried out over a time T_p. At the end of each period T_p, a pulse is transmitted from the unit A to the input of the unit B, which is proportional to the size of this average. It will further be shown that this average can be (with an error of the magnitude of the interference) taken as an approximate measure of the deviation from the extremum. In the unit B, the main component of the control is developed – the function x_1, which is then added to the oscillation x_2, giving the servosystem law for the extremum

$$x = x_1 + x_2.$$

As in all control systems, the extremal system above is realized by a certain combination of mathematical operations which transform the function y (t) into the function x (t).

In the case we are considering, the extremal system can be described by the expression*

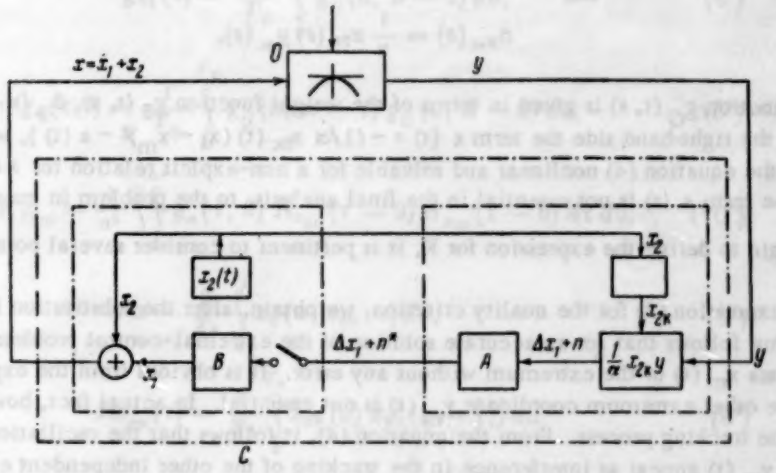
$$x(t) = \int_{-\infty}^{\infty} \left[g_B(t, s)\, \delta_B(s - t_p) \int_{-\infty}^{\infty} g_A(s, \theta) \frac{1}{\alpha} x_{2K}(\theta)\, y(\theta)\, d\theta \right] ds + x_2(t). \tag{1}$$

Here $x_2(t)$ and $x_{2K}(t)$ are symmetrical oscillations with period T_2 that have, in general, arbitrary forms except for the fact that for all values of t the inequality $x_2(t)\, x_{2K}(t) \geq 0$ holds. The ratio $T_p / T_2 = z$ is an integer, where T_p is the time over which the product $1/\alpha\, x_{2K}(t)\, y(t)$ is averaged. The coefficient α will be specified below. The weight function is

$$g_A(s, \theta) = \begin{cases} \dfrac{1}{T_p}, & s - T_p \leqslant \theta \leqslant s, \\ 0, & s - T_p > \theta;\ \theta > s; \end{cases}$$

*In the present article, we use the same notation as was used in [2].

$g_B(t, s)$ is an arbitrary stationary weight function, $\delta_B(s - t_p)$ the Dirac delta-function, and the t_p are equally spaced instants of time with the constant interval T_p. In the expression (1), the quantity t_p is considered as a function of s, which is for $s = t_p$ equal to t_p and for other values of s equal to zero.



The extremal system of the chosen form is therefore determined by the three arbitrary functions $g_B(t, s)$, $x_2(t)$, $x_{2K}(t)$, and the number T_p.

The object of control (see the sketch) is assumed to be inertialess and to have a static characteristic which is given by the displaced parabola*

$$y = y_m - (x - x_m)^2, \tag{2}$$

where x and y are the input and output coordinates of the object, and $x_m(t)$ and $y_m(t)$ are the coordinates of the extremal point of the characteristic. We will assume that $x_m(t)$ and $y_m(t)$ are stationary random function of the time.

As a criterion of the quality of control of the object, and consequently as a criterion of the quality of the extremal system, we use the expression**

$$K = \overline{y_m - y} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [y_m(t) - y(t)]\, dt. \tag{3}$$

This quality criterion is the error in following the extremum. The smaller the value of K, the better the quality of the extremal system.

The problem consists of finding the optimal characteristics of the extremal system that will ensure that K is a minimum***.

With this object in view, we derive below an analytical expression for the value of the quality criterion K in terms of the given characteristics $x_m(t)$ and $y_m(t)$ of an object with the given structure (2) and in terms of the arbitrary characteristics $g_B(t, s)$, T_p, $x_2(t)$ and $x_{2K}(t)$ of the extremal system (1).

2. The Equation of Motion of the System

In order to obtain an analytic expression for the tracking error, we need the equation of motion of the system OC (made up of the object O and the extremal system C), and its solution x(t) in explicit form in terms of $x_m(t)$,

*The constant in the equation of the parabola is taken to be unity, as this can always be brought by introducing a new variable x.

** Note that we always have $y_m - y \geq 0$.

*** The formulation of the problem and the method of solution we have indicated were communicated at the conference on the theory and application of discrete automatic systems held on September 26, 1958.

$y_m(t)$, and $g_B(t, s)$, T_p, $x_2(t)$, and $x_{2K}(t)$. The equation is (for the derivation of this equation, see the Appendix to this article)

$$x(t) = \int_{-\infty}^{\infty} g_G(t, s)\,[x_m(s) + n_{y_m}(s) + \varepsilon(s)]\,ds + x_2(t), \tag{4}$$

$$n_{y_m}(s) = \frac{1}{\alpha} x_{2K}(s)\, y_m(s),$$

where the weight function $g_G(t, s)$ is given in terms of the weight function $g_B(t, s)$, $\delta_B(s - t_p)$, and $g_A(s, \theta)$. The equation (4) has on the right-hand side the term $\epsilon(t) = -(1/\alpha\, x_{2K}(t)\,(x_1 - x_m)^2 - a(t))$, which is nonlinear relative to x_1. This makes the equation (4) nonlinear and solvable for a non-explcit relation for $x(t)$. It will be clear from what follows that the term $\epsilon(s)$ is not essential, in the final analysis, to the problem in question.

Before we begin to derive the expression for K, it is pertinent to consider several points concerning the equation (4).

A. From the expression (3) for the quality criterion, we obtain, after the substitution in it of (2), the result $K = (x - x_m)^2$. It thus follows that for an accurate solution of the extremal-control problem it is sufficient to follow one of the coordinates $x_m(t)$ of the extremum without any error. It is obvious from the expression (3) that the presence of oscillations in the other extremum coordinate $y_m(t)$ is not essential. In actual fact, however, these oscillations have an effect on the tracking process. From the equation (4), it follows that the oscillations of one independent extremum coordinate $y_m(t)$ appear as interference in the tracking of the other independent coordinate $x_m(t)$ of the extremal system. Acutally, to the function $x_m(t)$ which is to be tracked, we must add the component $n_{y_m}(t) = 1/\alpha\, x_{2K}(t)\, y_m(t)$ containing the oscillations of $y_m(t)$.

The oscillations of $y_m(t)$ appear as interference, independently of whether they are random or regular.

B. We note, relative to $y_m(t)$, that two separate components of the variation of the coordinate of the extremum $y_m(t)$ are usually considered: the "high-frequency noise", and the "slow creep". It is evident from the expressions (2) and (4), however, that both these components can be considered simultaneously.

C. From the equation (4), we note one peculilarity of the extremal system compared with a tracking system. The equation of motion, written in the form (4), shows that for extremal control we must track $x_m(t)$ after filtering out the interference $n_{y_m}(t)$. But the value of $x_m(t)$ cannot be measured. An extremal system thus has a relatively complicated measuring device for measuring the mismatch $\Delta x_1 = x_1 - x_m$. It is described at the beginning of section 1 and in [1], and is a correlator with a finite time of correlation. This device measures the derivative of the measured mismatch.

In ordinary servosystems, a subtractor is sufficient to measure the mismatch.

The differences we have described, that necessitate the use of different devices for measuring the mismatch, determine the features of optimal-extremum-system theory that differ from those of optimal-servosystem theory. In the theory of optimal-extremal systems, a more important place is occupied by the determination of the optimal characteristics of the measuring apparatus.

3. The Expression for the Quality Criterion K

The analytical expression for $K = K(x_m, y_m; g_B, T_p, x_2, x_{2K})$ is derived in the Appendix from the relations (3), (2), and (4). The final expression for K is most conveniently given as a chain of successive relations:

$$K = K_1 + K_2 = K_{1x_m} + K_{1y_m} + K_2 + k, \tag{11*}$$

$$K_2 = \overline{x_2^2} = \frac{1}{T_2} \int_0^{T_2} x_2^2(\theta)\, d\theta, \tag{1*}$$

$$K_{1x_m} = 2 \int_{-\infty}^{\infty} g_{01}(\tau)\, \Delta R_{x_m}(\tau)\, d\tau - \iint_{-\infty}^{\infty} g_0(\tau, \theta)\, \Delta R_{x_m}(\tau - \theta)\, d\tau\, d\theta, \tag{7*}$$

$$\Delta R_{x_m}(\tau) = R_{x_m}(0) - R_{x_m}(\tau_1),\ \tau_1 = \tau - \theta, \qquad (6^*)$$

$$g_{01}(\tau) = \frac{1}{T_p}\int_0^{T_p} g_G(u,\ u-\tau)\,du, \qquad (4^*)$$

$$g_0(\tau,\theta) = \frac{1}{T_p}\int_0^{T_p} g_G(u,\ u-\tau)\,g_G(u,\ u-\theta)\,du, \qquad (5^*)$$

$$K_{1y_m} = \frac{1}{\alpha^2}\int_{-\infty}^{\infty}\int g_0(\tau,\ \theta)\,R_{x_{2K}}(\tau-\theta)\,R_{y_m}(\tau-\theta)\,d\tau\,d\theta, \qquad (10^*) \qquad (5)$$

$$\alpha = \frac{2}{T_2}\int_0^{T_2} x_2(u)\,x_{2K}(u)\,du, \qquad (8^*)$$

$$R_{x_{2K}}(\tau) = \frac{1}{T_2}\int_0^{T_2} x_{2K}(u)\,x_{2K}(u-\tau_1)\,du. \qquad (9^*)$$

Here k is a nonlinear term related to ϵ (t), and it can be shown in our problem that it is relatively small.

The functions $R_{x_m}(\tau_1)$ and $R_{y_m}(\tau_1)$ are correlation functions for the laws of variation of the coordinates of the extremum $x_m(t)$ and $y_m(t)$.

The weight frunction $g_G(t, s)$ is determined from the system of equations

$$\left.\begin{aligned} g_V(t,\ s) &= \sum_{-\infty}^{\infty} g_B(t,\ t_p)\,g_A(t_p,\ s), && (2^*)\\ g_G(t,\ s) &= g_V(t,\ s) - \int_{-\infty}^{\infty} g_G(t,\ u)\,g_V(u,\ s)\,du. && (3^*) \end{aligned}\right\} \qquad (5)$$

The calculation of the value of K for given characteristics g_B, T_p, x_2, and x_{2K} of the extremal system (1) is performed in the order indicated by the numbers (1•) – (11•).

We now comment on certain points relative to the expression (5).

A. The following conclusion can be drawn from the expression for K: The"simplicity" (or "complexity") of the system OC is such that, for the calculation of K and for the derivation of the optimal characteristics of the extremal system, it is not necessary to know the laws $x_m(t)$ and $y_m(t)$ themselves. It is sufficient to know their correlation functions, i.e.,the linear extremal system that we are considering "varies" with the functions $x_m(t)$ and $y_m(t)$ only relative to their correlation functions $R_{x_m}(\tau)$ and $R_{y_m}(\tau)$.

A similar conclusion is obtained in the theory of optimal servosystems for any linear system.

In addition to this, the quantity K and the optimal characteristics of the system C are not directly dependent on $R_{x_m}(0)$ but depend only on ΔR_{x_m}, that is,on the form in practice of the correlation function in the neighborhood of $\tau_1 = 0$. In other words, K is independent of the "amplitude" of variation of the position of the extremum $x_m(t)$, and it depends on the "smoothness" of its variation.

This result permits a significant reduction in the extent of the characteristics that must be obtained from the object for the optimal adjustment in the case of a linear extremal system.

B. The component errors of tracking due to the variations of x_2, x_m, and y_m are independent. This property of the system OC permits a great simplification in the experimental and analytical investigation of an extremal system, and also in the representation of the final results of investigating the dependence of the size of the minimum tracking

error K_{min} on the tracking conditions. Such a possibility only occurs in the case of a parabolic characteristic. We note that in the case of a parabolic characteristic, K_2 is independent of Δx_1 and K_1.

C. In the case of a parabolic characteristic of the object, the quality criterion K in (3) reduces to the mean square error. This mathematically convenient quality criterion accurately reflects the technical requirements for the quality of an extremal system in the case when the output coordinate of the object satisfies the following condition: its supplementary decrease is equivalent for any y. This condition is satisfied by output coordinates such as efficiency coordinates, net cost, etc. If the output coordinate of the object is of another nature, then it can happen that the mean-square-error criterion will only approximately reflect the technical requirements. Such a situation occurs, for example, in radiolocation systems. It is known, however, from the theory of optimal servomechanisms, that the same system can be optimal for various quality criteria [2].

APPENDIX

1. Derivation of the Equation of Motion for the System OC (4)

For the sake of brevity, we introduce the linear-operation symbols

$$Az(t) = \int_{-\infty}^{\infty} g_A(s, \theta)\, z(\theta)\, d\theta,$$

$$Bz(t) = \int_{-\infty}^{\infty} g_B(t, s)\, \delta_B(s - t_p)\, z(s)\, ds.$$

Then the extremal system (1) can be given in the form

$$x = BA\left[\frac{1}{\alpha} x_{2K} y\right] + x_2 \quad \text{or} \quad x = C[y, c], \tag{1.1}$$

where C is a nonhomogeneous, linear, nonstationary operator, transforming the function y(t) into the function x(t), and independent of the arbitrary characteristics $c = g_B(t,s)$, T_p, $x_2(t)$, and $x_{2K}(t)$.

We must now obtain (4), i.e., we must express x(t) in terms of the functions $x_m(t)$, $y_m(t)$, and c.

From the general expression for $x = x_1 + x_2$, we see that we must obtain an expression for x_1 only, since the component x_2 is at our disposal, and is therefore known.

From the notation, we have $\Delta x = x - x_m = x_1 + x_2 - x_m = (x_1 - x_m) + x_2 = \Delta x_1 + x_2$, where $\Delta x_1 = x_1 - x_m$.

We make a preliminary transformation of the expression for y:

$$y = y_m - \Delta x^2 = y_m - (\Delta x_1 + x_2)^2 = -2\Delta x_1 x_2 + (y_m - \Delta x_1^2 - x_2^2). \tag{1.2}$$

We now substitute (1.2) into (1.1) and consider the transformation of A only:

$$A\left[\frac{1}{\alpha} x_{2K} y\right] = A\left[-\frac{1}{\alpha} 2\Delta x_1 x_2 x_{2K}\right] + A\left[\frac{1}{\alpha} x_{2K}(y_m - \Delta x_1^2 - x_2^2)\right]. \tag{1.3}$$

In the first term of (1.3) we break up the operator A, and divide the interval of integration into subintervals:

$$A\left[-\frac{1}{\alpha} 2\Delta x_1 x_2 x_{2K}\right] = -\frac{2}{\alpha T_p} \int_{t_p}^{t_p + T_p} \Delta x_1 x_2 x_{2K} d\theta = -\frac{2}{\alpha T_p} \sum_{i=1}^{z} \int_{T_{2i}} \Delta x_1 x_2 x_{2K} d\theta. \tag{1.4}$$

The mean value theorem for integrals is now applied to the last expression in (1.4); the conditions of applicability of this theorem are satisfied, since $x_2(t)\, x_{2K}(t) \geq 0$ for any t. We denote the point between the limits of integration in this theorem by θ_{iBH}. Then the right hand side of the expression (1.4) can be written in the form

$$-\frac{2}{\alpha T_p}\sum_{i=1}^{z}\Delta x_1(\theta_{iвн})\int_{T_{2i}} x_2 x_{2к}\,d\theta. \tag{1.5}$$

The integral in this expression is independent of i, and so we can reverse the order of integration and summation. The expression (1.5) becomes

$$-\frac{2}{\alpha T_p}\left(\int_0^{T_2} x_2 x_{2к}\,d\theta\right)\sum_{i=1}^{z}\Delta x_1(\theta_{iвн}). \tag{1.6}$$

On multiplying and dividing (1.6) by T_2, we now obtain

$$-\frac{2}{\alpha T_p T_2}\left(\int_0^{T_2} x_2 x_{2к}\,d\theta\right)\sum_{i=1}^{z}\Delta x_1(\theta_{iвн})\,T_2. \tag{1.6'}$$

We now define the arbitrary coefficient α by

$$\alpha=\frac{2}{T_2}\int_0^{T_2} x_2 x_{2к}\,d\theta,$$

and the expression (1.6') takes the form

$$-\frac{1}{T_p}\sum_{i=1}^{z}\Delta x_1(\theta_{iвн})\,T_2=-\frac{1}{T_p}\sum_{i=1}^{z}\Delta x_1(\theta_{iср})\,T_2-\frac{1}{T_p}\sum_{i=1}^{z}a_i T_2. \tag{1.7}$$

In (1.7) we use the notation $a_i=\Delta x_1(\theta_{iВН})-\Delta x_1(\theta_{iср})$ where $\theta_{iср}$ is the point of the period T_{2i}, where Δx_1 is equal to the mean value $\Delta x_{1ср}$ for this period.

We now transform the first sum in (1.7). Using the fact that $\Delta x_{1ср}=\frac{1}{T_2}\int_0^{T_2}\Delta x_1(\theta)\,d\theta$ and taking into account the structure of the operator A, we obtain

$$-\frac{1}{T_p}\sum_{i=1}^{z}\Delta x_1(\theta_{iср})\,T_2=-\frac{1}{T_p}\sum_{i=1}^{z}\int_{T_{2i}}\Delta x_1(\theta)\,d\theta=-\frac{1}{T_p}\int_0^{T_p}\Delta x_1(\theta)\,d\theta=-A[\Delta x_1]. \tag{1.8}$$

The second term in (1.7) yields

$$-\frac{1}{T_p}\sum_{i=1}^{z}a_i T_2=-A[a]. \tag{1.9}$$

The expression for $\underline{a}$ is derived in Section 7 of this Appendix. We consider the term $\underline{a}$ separately from the main part because in our problem it is, in the final analysis, not essential.

When we use (1.8) and (1.9) and the fact that $A\left[\frac{1}{\alpha}x_{2к}x_2^2\right]=0$, the expression (1.3) can be written as

$$A\left[\frac{1}{\alpha}x_{2к}y\right]=A[-\Delta x_1+n], \tag{1.10}$$

where $n=1/\alpha\ x_{2к}y_m-1/\alpha\ x_{2к}\Delta x_1^2+a=n_{y_m}+\epsilon$ [see (4)].

Corresponding to (1.1), we transform (1.10) by using the operator B, and obtain

$$x_1 = V[-\Delta x_1 + n], \tag{1.11}$$

where V = BA and V is thus a linear operator. On the basis of the fact that $\Delta x_1 = x_1 - x_m$, we make a series of obvious transformations of Eq. (1.11) and obtain

$$Ix_1 = V[-x_1 + x_m + n], \quad (I + V)x_1 = V[x_m + n] \tag{1.12}$$

where I is the identity operator.

We now introduce the operator Q^{-1} as the inverse of (I + V), which consequently satisfies the equation

$$Q^{-1}(I + V) = I, \tag{1.13}$$

Then from (1.12) we obtain

$$x_1 = G[x_m + n], \tag{1.14}$$

where $G = (I + V)^{-1}V$.

The equation of motion (4) of the system OC is now obtained from (1.1) and (1.14)

$$x = G[x_m + n] + x_2. \tag{1.15}$$

Certain comments follow on the derivation of the equation of motion (4).

A. We first of all explain the physical significance of the operator $A\left[\frac{1}{\alpha}x_{2K}y\right]$, which determines the deviation from the extremum. Firstly, it determines the mean over the time T_p of the value of the derivative dy/dx at points of the objective characteristic removed from the extremum by $\Delta x_1(t)$, where t varies from 0 to T_p. Besides this, the mean value of n(t) is obtained as a result of measurement. The quantity Δx_1 is unknown, and is, in fact, an object of the measurement. The derivative dy/dx has a functional relationship with Δx_1. The mean value of dy/dx can thus determine a certain mean value of Δx_1. In this connection, the quantity Δx_1 appears in the results of applying the transformation A (1.10). Since dy/dx depends linearly on Δx_1 for a parabola, the quantity Δx_1 occurs to the first power.

B. The expression $1/\alpha\, x_{2K}\, y$, when the transformation A is applied to it, is written in the form $\Delta x_1(t) + n(t)$, where n(t) is the interference. After applying the transformation A, we obtain another function, which is the output of the unit A (see the description of the system in Section 1), which can also be written in the form $\Delta x_1(t) + n^*(t)$. Since the extremal system is a pulse system, there is a sense in which we can consider the output of the unit A at the instant of time t_p only. For the time t_p, the interference $n^*(t)$ has the form $n^*(t_p) = \overline{\Delta x_1(t)} - \Delta x_1(t_p) + \overline{n(t)}$, where the bar denotes, as usual, the mean of a quantity over the time T_p. Consequently, at the output of the unit A, the error in measuring the quantity $\Delta x_1(t_p)$ is related first of all to the measurement mean, and second to the presence of the component n(t).

For the method of application in the present paper, the error arising in the unit A is not separated out; it is formed by the general linear transformation G.

C. The expression (1.10) can easily be obtained if we assume that the fluctuations in x_2 and x_{2K} are rectangular; in this case the coefficient of Δx_1 in the expression (1.3) is unity. We will consider a more general case than this, in order to be able to find the optimal form of the variation of x_2 and x_{2K}.

D. In practice, the derivative dy/dx is calculated by the appartus A in the form of the incremental ratio $\Delta y/\Delta x$. In the case of a parabola, we have dy/dx independent of the amplitude of the test oscillations of x_2.

E. We note that the oscillations of x_2 and x_{2K} can also be random.

2. Derivation of the Equations (2*) and (3*) Given in (5), Relating the Weight Function $g_G(t,s)$ to $g_A(t,s)$ and $g_B(t,s)\,\delta_B(\delta - t_p)$

We can show that the weight function for the operator V = BA is given by the expression

$$g_V(t, \theta) = \sum_{t_p=-\infty}^{\infty} g_B(t, t_p)\, g_A(t_p, \theta).$$

It remains to show that if the operators V and Q^{-1} are given [Eq. (1.13)], then the weight function of the operator $G = Q^{-1}V$ is determined by the integral equation (3•) of the system (5).

In operational form, the equation (3•) is obtained from $G = (1 + V)^{-1}V$ as follows:

$$(1 + V)\,G = (1 + V)(1 + V)^{-1}V,\ (1 + V)\,G = V,\ G + VG = V,\ G = V - VG.$$

When we write this relation out in detail, we obtain (3•).

3. The Derivation of $K = K(x_m, y_m, c)$

From (3) and (2), we obtain

$$K = \overline{y_m - y} = \overline{\Delta x^2},\ \Delta x = x - x_m,$$

i.e., in our case, the quality criterion K reduces to the mean-square-error criterion. If we use the fact that $\Delta x = \Delta x_1 + x_2$, and assume for simplicity that the mean-time $\Delta x_1 x_2 = 0$, we obtain

$$K = K_1 + K_2,$$

where $K_1 = \overline{\Delta x_1^2}$ is the component of the tracking error corresponding to the tracking error that would occur in an ordinary servomechanism it if were tracking $x_m(t)$ in the presence of interference, and $K_2 = \overline{x_2^2} = \frac{1}{T_2}\int_0^{T_2} x_2^2(t)\,dt$ is the component of the tracking error K due to the supplementary oscillatory motion of x_2 that is necessary in carrying out the measurement of $\Delta x_1(t)$. The component K_2 is thus already expressed in the terms of the characteristic of the system OC. It remains to obtain the expression for the component K_1. If we use the fact that $\Delta x_1 = x_1 - x_m$, we have

$$K_1 = \overline{(x_1 - x_m)^2}$$

We now use the equation of motion of the system (1.15). If we assume for simplicity that the mean-time $\overline{x_m y_m} = 0$, we obtain

$$K_1 = K_{1x_m} + K_{1y_m} + k,$$

where

$$K_{1x_m} = \overline{(G_1 x_m)^2},\ G_1 = G - I, \tag{3.1}$$

$$K_{1y_m} = \overline{(Gn_{y_m})^2},\ n_{y_m} = 1/\alpha\ x_{2\kappa}\, y_m. \tag{3.2}$$

The term k is related to the function $\epsilon(t)$. In the second part of this article we will estimate its size, and it will be shown that in our problem it tends to zero for an optimal extremal system

In order to simplify the form of subsequent transformations, we will introduce them on the basis of identities which are first derived in other parts of this appendix. We now give the transformation scheme.

If we use the identity (4.2), then from (3.1) and (3.2) we obtain

$$K_{1x_m} = G_{I_1} R_{x_m}(\tau_1), \tag{3.3}$$

$$K_{1y_m} = G_{0_1} R_{n_{y_m}}(\tau_1), \tag{3.4}$$

where G_{I_1} and G_{01} are linear operators. Thus, the component of the tracking error due to the variation of the position x_m and the size y_m of the extremum is obtained by a linear transformation of the correlation functions R_{x_m} and $R_{n_{y_m}}$. If we use the identity (5.1) in (3.3), we obtain the final expression for K_{1x_m} in the form (7•) of the system (5). By using the identity (6.1) in the expression (3.4), we transform the correlation function $R_{n_{y_m}}(\tau_1)$, and the error component K_{1y_m} is obtain in the form (10•) in (5).

4. The Derivation of the Identity $\overline{(G_1 z(t))^2} = G_{I_1} R_z(\tau_1)$

We carry out this derivation using the notation corresponding to that used in the transition from (3.1) to 3.3).

We note that the identity (4.2) is also used in obtaining (3.4). In this case, we also use the correspondence in notation $G_I \sim G$ and $G_{I_1} \sim G_{01}$.

It is therefore necessary to prove the following.

If the nonstationary part of the weight function $g_I(t, s)$ of the integral operator G_I is periodic with period T_p; i.e., if

$$g_I(t, s) = g_I(t + T_p, s + T_p), \tag{4.1}$$

and if, in addition to this, $z(t)$ is a stationary random process with correlation function $R_z(\tau_1)$, then

$$\overline{(G_I z(t))^2} = G_{I_1} R_z(\tau_1), \tag{4.2}$$

$$G_{I_1}[u(\tau_1)] = \int_{-\infty}^{\infty}\int g_{I_1}(\theta, \tau)\, u(\theta - \tau)\, d\tau d\theta \tag{4.3}$$

is a linear operator, where

$$g_{I_1}(\theta, \tau) = \frac{1}{T_p}\int_0^{T_p} g_I(s, s - \theta)\, g_I(s, s - \tau)\, ds, \tag{4.4}$$

with $\tau_1 = \theta - \tau_1$ and $R_z(\tau_1)$ the correlation function.

For the proof, we write the left hand side of (4.2) in the form $\overline{G_I z(t)\, G_I z(t)}$ and consider the product

$$G z(t)\, G_I z(t) = \Big[\int_{-\infty}^{\infty} g(t, r)\, z(r)\, dr\Big]\Big[\int_{-\infty}^{\infty} g_I(t, u)\, z(u)\, du\Big] =$$

$$= \int_{-\infty}^{\infty}\int dr\, du\, g_I(t, r)\, g_I(t, u)\, z(u)\, z(r).$$

If we use the fact that the averaging operation with respect to time can be introduced under the integration sign, we obtain

$$\int_{-\infty}^{\infty}\int dr\, du \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} g_I(t, r)\, g_I(t, u)\, z(u)\, z(r)\, dt. \tag{4.5}$$

We fix r and u relative to the time t, and set

$$u = t - \tau, \qquad r = t - \theta, \tag{4.6}$$

where τ and θ are fixed quantities. Thus, u and r are variables in the integration with respect to t. After this change in variables, (4.5) takes the form

$$\int_{-\infty}^{\infty}\int d\tau d\theta \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} g_I(t, t - \theta)\, g_I(t, t - \tau)\, z(t - \tau)\, z(t - \theta)\, dt. \tag{4.7}$$

The change in variables (4.6) does not alter the value of the double integral, as can be verified by calculating the corresponding Jacobian.

When the variables are related by (4.6), then from (4.1) we have

$$g_I(t, t - \tau)\, g_I(t, t - \theta) = g_I(t + T_p, t + T_p - \tau)\, g_I(t + T_p, t + T_p - \theta). \tag{4.8}$$

We consider the integrand of the inside integral in the expression (4.7) at the times t_p^*, with distance between them of T_p, and which are displaced relative to the times t_p by the fixed time t^*: thus, $t_p^* = t_p + t^*$.

The integrand in (4.7) thus possesses the property that at the times t_p^*, the factor $g_I(t_p^*, t_p^* - \tau)\, g_I(t_p^*, t_p^* - \theta)$ takes identical values. When the times t_p^* change (for a change in t^*), then the value of this factor also changes in general.

Taking into account what has been said above, we transform the internal integral in (4.7)

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g_I(t, t-\tau)\, g_I(t, t-\theta)\, z(t-\tau)\, z(t-\theta)\, dt. \tag{4.9}$$

We divide up the range of integration (T, – T) into subintervals, set $T/T_p = N$, and obtain

$$\begin{aligned}
&\lim_{T\to\infty} \frac{1}{2T} \sum_{p=-N}^{N} \int_{t_p}^{t_{p+1}} g_I(t, t-\tau)\, g_I(t, t-\theta)\, z(t-\tau)\, z(t-\theta)\, dt \cdot \\
&= \lim_{T\to\infty} \frac{1}{2T} \sum_{p=-N}^{N} \int_{0}^{T_p} g_I(t_p+t', t_p+t'-\tau)\, g_I(t_p+t', t_p+t'-\theta)\, z(t_p+t'-\tau) z(t_p+t'-\theta)\, dt' \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{0}^{T_p} \left\{ \sum_{p=-N}^{N} g_I(t_p+t', t_p+t'-\tau) g_I(t_p+t', t_p+t'-\theta) z(t_p+t'-\tau) z(t_p+t'-\theta) \right\} dt'.
\end{aligned} \tag{4.10}$$

Here we have used the fact that $t_p^* = t_p + t^*$, and that we can interchange the order of summation and integration. As we showed above, when t^* is fixed, the factor (4.8) is independent of p, and so in (4.10) it can be taken out from under the sign of summation with respect to p, and the index p in it suppressed, for example, by setting $t_p = t_{p=0} = 0$. We obtain for the right hand side of (4.10) the expression

$$\lim_{T\to\infty} \frac{1}{2T} \int_{0}^{T_p} \left[g_I(t', t'-\tau)\, g_I(t', t'-\theta) \sum_{p=-N}^{N} z(t_p + t' - \tau)\, z(t_p + t' - \theta) \right] dt'. \tag{4.11}$$

When we now take into account the fact that the limit with respect to T refers only to the sum in the integrand, (4.11) can be written as

$$\int_{0}^{T_p} \left\{ g_I(t', t'-\tau)\, g_I(t', t'-\theta) \lim_{T\to\infty} \frac{1}{2T} \sum_{p=-N}^{N} z(t_p + t' - \tau)\, z(t_p + t' - \theta) \right\} dt'. \tag{4.12}$$

If we now refer to the fact that the process z (t) is assumed to be stationary, we see that the limit of the sum in (4.12) is independent of the time displacements $t_p^* = t_p + t^*$, i.e., it is not dependent on the value of t^*. The summation with respect to p can therefore be taken out from under the sign of integration with respect to t^* as a constant factor, and we can subsequently set $t^* = 0$. The expression (4.12) now takes the form

$$\left[\int_{0}^{T_p} g_I(t', t'-\tau)\, g_I(t', t'-\theta)\, dt' \right] \left[\lim_{T\to\infty} \frac{1}{2T} \sum_{p=-N}^{N} z(t_p - \tau)\, z(t_p - \theta) \right]. \tag{4.13}$$

We now change the notation for the summation variable in (4.13); we replace $t_p - \tau$ by t_p, and then $t_p - \theta$ will become $t_p + \tau - \theta$, and the sum in (4.13) becomes

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{p=-N}^{N} z(t_p)\, z(t_p + \tau_1),$$

where

$$\tau - \theta = \tau_1$$

It is known [3] that

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{p=-N_*}^{N_*} z(t_p)\, z(t_p + \tau_1) = \frac{1}{T_p} R_z(\tau_1). \tag{4.14}$$

If we use this result, the expression (4.5) can be written

$$\int_{-\infty}^{\infty}\int d\tau\, d\theta\, g_{I1}(\tau, \theta)\, R_z(\tau - \theta),$$

where we have used the notation of (4.4). We finally denote the operation (4.3) by G_{I1}, and obtain the identity (4.2).

5. The Derivation of the Identity

$$G_{I1}R_{x_m}(\tau_1) = 2\int_{-\infty}^{\infty} g_{01}(\tau)\,\Delta R_{x_m}(\tau)\,d\tau - \int_{-\infty}^{\infty}\int g_0(\tau,\theta)\,\Delta R_{x_m}(\tau-\theta)\,d\tau\,d\theta$$

In this identity, we use the notation of the system (5)[see (4•), (5•), and (8•)], and also set $G_I = G - I$ [see (3.1)].

For the proof of this identity, we write the weight function of the operator $G_I = G - I$ as

$$g_I(t, s) = g_G(t, s) - \delta(t - s), \tag{5.1}$$

which satisfies the condition (4.1). When we use (5.1), then from (4.4) we obtain the expression for the weight function $g_{I1}(\tau, \theta)$:

$$\begin{aligned} g_{I1}(\tau, \theta) &= \frac{1}{T_p}\int_0^{T_p} g_I(u, u-\theta)\, g_I(u, u-\tau)\, du \\ &= \delta(\theta)\,\delta(\tau) - \delta(\tau)\, g_{01}(\theta) - \delta(\theta)\, g_{01}(\tau) + g_0(\theta, \tau), \end{aligned} \tag{5.2}$$

where we have used the notation (4•) and (5•).

We now use the operator G_{I_1} to transform the correlation function $R_{x_m}(\tau_1)$, and obtain

$$G_{I1}R_{x_m}(\tau_1) = \int_{-\infty}^{\infty}\int g_{I1}(\tau, \theta)\, R_{x_m}(\tau - \theta)\, d\tau d\theta. \tag{5.3}$$

The substitution of (5.2) in (5.3) yields

$$R_{x_m}(0) - \int_{-\infty}^{\infty} g_{01}(\theta)\, R_{x_m}(-\theta)\, d\theta - \int_{-\infty}^{\infty} g_{01}(\tau)\, R_{x_m}(\tau)\, d\tau + \int_{-\infty}^{\infty}\int g_0(\tau, \theta)\, R_{x_m}(\tau - \theta)\, d\tau d\theta. \tag{5.4}$$

When we use the fact that $R_{x_m}(-\theta) = R_{x_m}(\theta)$, and change the notation for the variable of integration in the first integral in (5.4), we obtain

$$R_{x_m}(0) - 2\int_{-\infty}^{\infty} g_{01}(\tau)\, R_{x_m}(\tau)\, d\tau + \int_{-\infty}^{\infty}\int g_0(\tau, \theta)\, R_{x_m}(\tau - \theta)\, d\tau d\theta. \tag{5.5}$$

We further transform (5.5) by using the relation $R_{x_m}(\tau_1) = R_{x_m}(0) - \Delta R_{x_m}(\tau_1)$ [see (8•) of the system (5)] and the identities

$$\int_{-\infty}^{\infty} g_{01}(\tau) = 1, \quad \iint_{-\infty}^{\infty} g_0(\tau, \theta)\, d\tau d\theta = 1.$$

Then the single and double integrals in (5.5) become respectively

$$R_{x_m}(0) - \int_{-\infty}^{\infty} g_{01}(\tau)\, \Delta R_{x_m}(\tau)\, d\tau, \tag{5.6}$$

$$R_{x_m}(0) - \iint_{-\infty}^{\infty} g_0(\tau, \theta)\, \Delta R_{x_m}(\tau - \theta)\, d\tau d\theta. \tag{5.7}$$

When we substitute (5.6) and (5.7) in (5.5), we finally obtain the desired identity.

Proof of the Identity $R_{n_{y_m}}(\tau_1) = (1/\alpha^2) R_{x_{2k}}(\tau_1) R_{y_m}(\tau_1)$

In the expression for the identity, the interference n_{y_m} has the form $n_{y_m} = 1/\alpha\, x_{2K} y_m$. For the proof, we write the formula for the left-hand side

$$R_{n_{y_m}}(\tau_1) = \frac{1}{\alpha^2} \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_{2n}(t)\, y_m(t)\, x_{2n}(t - \tau_1)\, y_m(t - \tau_1)\, dt. \tag{6.1}$$

We first of all note a certain relation and introduce the notation. Since $x_{2k}(t)$ is a periodic function, the product $x_{2k}(g)x_{2k}(t-\tau_1)$ is also periodic. We denote this product, for fixed τ_1, by $x(t)$. Then

$$x(t) = x_{2K}(t)\, x_{2K}(t - \tau_1) \text{ and } x(t) = x(t + T_2). \tag{6.2}$$

We also use, for fixed τ_1, the notation

$$y(t) = y_m(t)\, y_m(t - \tau_1) \tag{6.3}$$

In the notation of (6.2) and (6.3), the expression for the limit in (6.1) becomes

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, y(t)\, dt. \tag{6.4}$$

The further transformations are similar to those in Section 4.

We divide up the range of integration of (6.4) and obtain

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{i=-N}^{N} \int_{t_i}^{t_{i+1}} x(t)\, y(t)\, dt, \tag{6.5}$$

where $N = T/T_2$ and $t_i = iT_2$.

The change of variables $t = t_i + t'$ is introduced, yielding

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{i=-N}^{N} \int_{0}^{T_2} x(t_i + t')\, g(t_i + t')\, dt' = \lim_{T\to\infty} \frac{1}{2T} \int_{0}^{T_2} \left[\sum_{i=-N}^{N} x(t_i + t')\, y(t_i + t') \right] dt' \tag{6.6}$$

For fixed t', the factor $x(t_i + t')$ of period T_2 is independent of i, and can be taken out from under the summation with respect to i, and we can then set $t_i = t_{i=0} = 0$ for simplicity. The right hand side of (6.6) can now be written

$$\lim_{T\to\infty}\frac{1}{2T}\int_0^{T_2}\left[x(t')\sum_{i=-N}^{N}y(t_i+t')\right]dt'. \tag{6.7}$$

Since the limit with respect to T refers only to the sum with respect to i, (6.7) can be written

$$\int_0^{T_2}\left[x(t')\lim_{T\to\infty}\frac{1}{2T}\sum_{i=-N}^{N}y(t_i+t')\right]dt' \tag{6.8}$$

The process y(t) has been assumed to be stationary, and so the limit of the sum is independent of the displacement of the time $(t_i + t')$, i.e., it is independent of t'. The limit of the sum can therefore be taken out from under the sign of integration as a constant factor, and we may then set $t' = 0$. for simplicity. When we also use the expression (6.1), we obtain

$$\frac{1}{\alpha^2}\left[\int_0^{T_2}x(t')\,dt'\right]\left[\lim_{T\to\infty}\frac{1}{2T}\sum_{i=-N}^{N}y(t_i)\right]. \tag{6.9}$$

The use of the notation in (6.3) and the identity (4.16) now yield

$$\lim_{T\to\infty}\frac{1}{2T}\sum_{i=-N}^{N}y(t_i)=\frac{1}{T_2}R_{y_{in}}(\tau_1). \tag{6.10}$$

If we now use (6.10), the notation of (6.2) and (9•) from the system (5), we obtain from (6.9) the right-hand side of the identity that was to be proved.

7. Derivation of the Expression for a [see (13)]

We now assume that, in a range including the period T_{2i} about its mean, we can expand Δx_1 in a Maclaurin series. With this assumption, we calculate $\Delta x_1(\theta_{icp})$ and $\Delta x_1(\theta_{iBH})$.

When only three terms of the series for $\Delta x_1(\theta_{icp})$ are retained, we have

$$\Delta x_1(\theta_{icp})=\Delta x_{icp}\approx b_0+\frac{1}{24}\left(\frac{d^2\Delta x_1(\theta)}{d\theta^2}\right)_{\theta=0}T_2^2. \tag{7.1}$$

Similarly, we have

$$\Delta x_1(\theta_{inn})=\frac{\int_{T_{2i}}\Delta x_1x_2x_{2n}d\theta}{\int_{T_{2i}}x_2x_{2n}d\theta}\approx b_0+\frac{1}{2}\left(\frac{d^2\Delta x_1(\theta)}{d\theta^2}\right)_{\theta=0}\frac{\int_{-T_2/2}^{T_2/2}\theta^2x_2x_{2n}d\theta}{\int_{-T_2/2}^{T_2/2}x_2x_{2n}d\theta}. \tag{7.2}$$

From (1.7), we obtain

$$a_i=\frac{T_2^2\psi}{24}\left(\frac{d^2\Delta x_1(\theta)}{d\theta^2}\right)_{\theta=0}, \tag{7.3}$$

where

$$\psi[x_2, x_{2H}] = -\left(1 - \frac{12\int_{-T_2/2}^{T_2/2} \theta^2 x_2 x_{2H}\, d\theta}{T_2^2 \int_{-T_2/2}^{T_2/2} x_2 x_{2H}\, d\theta}\right)$$

is a functional of $x_2 x_{2H}$; for rectangular oscillations the functional ψ is equal to zero, while for sinusoidal oscillations it is approximately equal to −0.15. We now consider the identity

$$\left(\frac{d^2 \Delta x_1(\theta)}{d\theta^2}\right)_{\theta=0} = \left(\frac{d^2 \Delta x_1(\theta)}{d\theta^2}\right)_{\text{mean over } T_{2i}} + \varepsilon_i = \frac{1}{T_2}\int_{T_{2i}} \left(\frac{d^2 \Delta x_1(\theta)}{d\theta^2}\right)_{\theta=t} dt + \varepsilon_i. \tag{7.4}$$

The quantity ϵ_i becomes smaller when the law of variation of the second derivative becomes closer to a linear law and when T_2 decreases. If the function $\Delta x_1(t)$ can be approximated sufficiently accurately by a polynomial of the third degree, then $\epsilon_i = 0$.

We use (7.4) to transform (7.3), and obtain the approximate result

$$a_i = \frac{T_2^2 \psi}{24} \frac{1}{T_2} \int_{T_{2i}} \left(\frac{d^2 \Delta x_1(\theta)}{d\theta^2}\right)_{\theta=t} dt.$$

The second term in (1.7) takes the form

$$-\frac{1}{T_p}\sum_{i=1}^{s} a_i T_1 = -\frac{T_2^2 \psi}{24}\frac{1}{T_p}\int_{t_p}^{t_p+T_p} \left(\frac{d^2 \Delta x_1(\theta)}{d\theta^2}\right)_{\theta=t} = -A[a],$$

$$a = \frac{T_2^2 \psi}{24}\left(\frac{d^2 \Delta x_1(\theta)}{d\theta^2}\right)_{\theta=t}.$$

LITERATURE CITED

1. Grishko, N. V., "Extremal control with extremum tracking", Avtomatika i telemekhanika, Vol. 20, No. 4, 1959.
2. Pugachev, V. S., The Theory of Random Functions and its Application to Problems of Automatic Control. Fizmatgiz, 1960.
3. Tsypkin, Ya. Z., "Thedesign of discontinuous control systems in the presence of stationary random disturbances." Avtomatika i telemekhanika, Vol. 14, No. 4, 1953.

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

ON THE QUESTION OF CHOICE OF PARAMETERS IN THE CONTROL PART OF A GRADIENT-TYPE AUTOMATIC OPTIMIZATION SYSTEM

T. I. Tovstykha (Moscow)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 1027-1037, August, 1961
Original article submitted July 25, 1960

The parameters of the control part of a gradient-type automatic optimization system with one or several independent inputs are chosen with the consideration that the steady value of the mathematical expectation of the output signal should be sufficiently small.

In the case of a system with one input the choice of operating parameters is made with consideration taken of limitations of the average spread about the steady value of the output signal. The rapidity of action of the system is estimated.

1. Formulation of the Problem

In Fig. 1 a schematic diagram of a discrete automatic optimization system with m independent inputs is presented, in which a search for the extremum (minimum) is performed by the method of gradients. The letter O denotes the control object, inertialess, and its characteristics are considered parabolic:

$$y[n] = \sum_{j=1}^{m} x_j^2[n]. \tag{1.1}$$

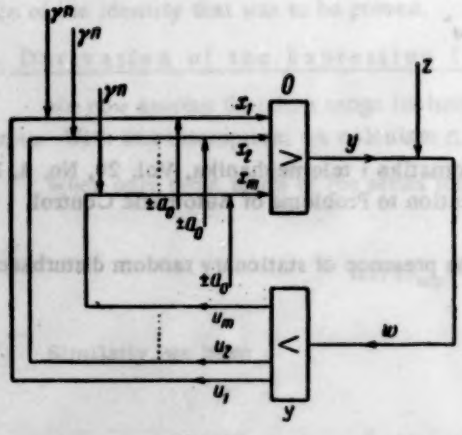


Fig. 1.

Here all quantities are considered at discrete moments of time, n is the number of the corresponding cycle, referring to the moment t = nT. We shall use L [n] to denote the value of the quantity L at the moment of the time nT.

A random disturbance Z [n] acts at the output. We suppose that the whole time duration of each cycle is divided into m intervals. We consider the i-th interval within the n-th cycle. Clearly the beginning of the interval corresponds to the moment (time scale is referred to the duration T of a cycle) $n - 1 + (i - 1)/m$, the center, to the moment $n - 1 + (i - \frac{1}{2})/m$, and the end, to the moment $n - 1 + i/m$. As regards the disturbances Z [n] we assume that its increments in time during a half interval (for definiteness, the i-th interval in the n-th cycle)

$$v_i[n-1] = z\left[n-1+\frac{i-\frac{1}{2}}{m}\right] - z\left[n-1+\frac{i}{m}\right] \qquad (i = 1, 2, \ldots, m) \tag{1.2}$$

are independent random quantities with probability density $\rho(v)$. In future calculation it will be required to know the moments v [n]. For definiteness we suppose that all odd moments are equal to zero, the dispersion is finite, and equal to σ^2, and also that $M(v^4) = 3\sigma^4$ and $M(v_i^2 v_j^2) = \sigma^4$, $i \neq j$. The remaining moments of fourth order we consider to be zero. The values of the moments taken coincide with the values of a normal distribution law [1].

It is assumed that a nonrandom disturbance $z_0[n]$, brought about by a displacement of the characteristics

$$z_0[n] = \gamma n, \tag{1.3}$$

is added to the signal from the control part γ in each of the m channels at the input to the system.

For simplicity we suppose the rate γ to be the same in all channels. One could also consider the case of differing rates, or even the case when the γ_i are random quantities.

We consider the n-th cycle. As was already indicated above, the entire time of the cycle we consider divided into m intervals. At the beginning of the i-th interval in the i-th input to the system a trial disturbance $+ a_0$ is given, and at the center of the interval, a disturbance $- a_0$. At the center and end of the interval, respectively, output quantities are measured, and their difference is applied at the control device. This difference is

$$\begin{aligned}\Delta w_i[n] &= w\left[n-1+\frac{i-\frac{1}{2}}{m}\right] - w\left[n-1+\frac{i}{m}\right] \\ &= y\left[n-1+\frac{i-\frac{1}{2}}{m}\right] + z\left[n-1+\frac{i-\frac{1}{2}}{m}\right] - y\left[n-1+\frac{i}{m}\right] \\ &\quad - z\left[n-1+\frac{i}{m}\right].\end{aligned} \tag{1.4}$$

The difference in values of the disturbances is clearly equal to the quantities $v_i[n-1]$ defined by (1.2). A signal $u_i[n]$ from the control part of the i-th input is applied to the beginning of the succeeding $(n+1)$ - th cycle; here the increment in the quantity u_i is proportional to the difference $\Delta\omega_i[n]$; i.e.,

$$u_i[n] - u_i[n-1] = \Delta u_i[n] = -a\Delta\omega_i[n]. \tag{1.5}$$

Here a is a characteristic coefficient.

The problem consists in determining the quantities a, a_0 corresponding to the admissible sufficiently small stable ($n \to \infty$) values of the mathematical expectation m_y of the output quantities y. From formula (1.1) it follows that in the problem considered

$$m_y = \sum_{i=1}^{m}(m_{x_i}^2 + D_{x_i}). \tag{1.6}$$

Thus, it is necessary to determine m_x and D_x for each of the input variables.

An analogous but more general problem is considered in [2]; the inputs are considered interdependent. The equations obtained, from which one may determine the limiting value as $n \to \infty$ of the arithmetical mean of the output quantities for various different tests. In the example for $m = 1$ a graphical dependence of this limiting value on quantities proportional to the product of the parameters a and a_0 is constructed.

In the present article the solution of a system of difference equations reduces the calculations immediately to concrete results allowing the choice of operating parameters a and a_0 in the case of any number of independent inputs. In the case of a single input, not only the dependence $m(a, a_0)$, but also the dependence of the dispersion D_y on those parameters a, a_0 is considered, which dependence allows one, through the choice of the parameters, to consider not only the steady magnitude m_y, but also the magnitude of the mean spread about m_y. In section 2 a possible estimate of the rapidity of action proposed by A. A. Fel'dbaum is investigated.

2. Case of a Single Input

In the case of one input ($m = 1$), at the beginning of the cycle a test $+ a_0$ is applied at the input, and at the middle of the cycle, a test $- a_0$. It is considered that the reaction of the object to the disturbance supplied may be measured at its output after a time $T/2$. In correspondance with (1.5) the difference of values of the output quantities, measured in the middle and at the end of the cycle, are proportional to the difference in controlling signals u given at the input of the object in the given cycle and in the preceeding. One may write that

We expand $\Delta u\,[n+1]$: $x\,[n+1] = x\,[n] + \gamma + \Delta u\,[n+1]$.

or $$\Delta u\,[n+1] = -a\Delta w\,[n+1] = -a\left\{(x\,[n]+a_0)^2 - \left(x\,[n] - a_0 + \frac{\gamma}{2}\right)^2 + v\,[n]\right\}$$

and $$\Delta u\,[n+1] = -a\left\{4\left(a_0 - \frac{\gamma}{4}\right)x\,[n] + \gamma\left(a_0 - \frac{\gamma}{4}\right) + v\,[n]\right\}$$

$$x\,[n+1] = x\,[n] + \gamma - a\left\{4\left(a_0 - \frac{\gamma}{4}\right)x\,[n] + \gamma\left(a_0 - \frac{\gamma}{4}\right) + v\,[n]\right\}. \tag{2.1}$$

Solving this difference equation, we obtain

$$x\,[n] = (1-4a\delta)^n\,x\,[0] + \sum_{s=0}^{n-1}(1-4a\delta)^s\{\gamma(1-a\delta) - av\,[n-s]\}, \tag{2.2}$$

where

$$\delta = a_0 - \frac{\gamma}{4}. \tag{2.3}$$

In this formula x [0] is the initial value of $\underline{x}$. We shall consider that the initial value is a random quantity with mathematical expectation $m_x\,[0]$ and dispersion $D_x\,[0]$.

Transforming from x [n] to $m_x\,[n]$ in (2.2) and taking into account that m (v) = 0, we easily obtain

$$m_x\,[n] = (1-4a\delta)^n\,m_x\,[0] + \frac{\gamma(1-a\delta)}{4a\delta}[1-(1-4a\delta)^n]. \tag{2.4}$$

Here aδ may take on any value from the open interval $[0, \frac{1}{2}]$; i.e., $|1-4a\delta| < 1$.

We now calculate the dispersion $D_x\,[n]$. It is clear that

$$D_x\,[n] = \overline{(x\,[n] - m_x\,[n])^2}$$
$$= \left\{(1-4a\delta)^n(x\,[0] - m_x\,[0]) - a\sum_{s=0}^{n-1}(1-4a\delta)^s v\,[n-s]\right\}^2.$$

Since the right side of (2.2) is a sum of uncorrelated random quantities ($\underline{v}$ is uncorrelated by assumption; x [0] and $\underline{v}$ may also be considered uncorrelated), the dispersion $D_x\,[n]$ is equal to the sum of the dispersions of these quantities; i.e.,

$$D_x\,[n] = \frac{\sigma^2}{8\delta^2}\frac{a\delta}{1-2a\delta}[1-(1-4a\delta)^{2n}] + (1-4a\delta)^{2n}D_x\,[0]. \tag{2.5}$$

Passing to the limit $n \to \infty$ in (2.5) and (2.4), we obtain the following expression for the steady values of the mathematical expectation and dispersion:

$$m_x = \frac{\gamma(1-a\delta)}{4a\delta}, \tag{2.6}$$

$$D_x = \frac{\sigma^2}{8\delta^2}\frac{a\delta}{1-2a\delta}. \tag{2.7}$$

In accordance with (1.6)

$$m_y = \gamma^2\left(\frac{1-a\delta}{4a\delta}\right)^2 + \frac{\sigma^2}{8\delta^2}\frac{a\delta}{1-2a\delta}. \tag{2.8}$$

To calculate the dispersion of the output quantities we use the relation

$$D_y = M[(y-m_y)^2] = M(x^4) - (M(x^2))^2$$

The quantity $M(x^2)$ is calculated in (2.8). We determine $M(x^4)$. From (2.2), denoting $v[n-s]$ by $v^{(s)}$, as $n \to \infty$ we have:

$$x = m_x - L, \quad L = a\sum_{s=0}^{\infty}(1-4a\delta)^s v^{(s)}.$$

From this,

$$x^4 = m_x^4 - 4m_x^3 L + 6m_x^2 L^2 - 4m_x L^3 + L^4.$$

Passing from the equation for x^4 to the corresponding equation for the moments, and discarding the moments of oddorder of L (i.e., ν), we obtain

$$M(x^4) = m_x^4 + 6m_x^2 M(L^2) + M(L^4).$$

It is easy to see that

$$M(L^2) = \frac{a^2\sigma^2}{1-(1-4a\delta)^2} = D_x.$$

In the transition to moments in the equation for $M(x^4)$ we first of all throw out the terms containing odd powers $v^{(s)}$, and also omit sums of products of the type $v^{(i)}v^{(j)}v^{(k)}v(l)$, $v^{(i)}v^{(j)}v^{(k)^2}$, $v^{(i)}v^{(j)^2}$, when all indices i, j, k, l are different (see the assumptions on the moments). As a result we obtain

$$M(L^4) = a^4 M\left\{\sum_{s=0}^{\infty}[(1-4a\delta)^s v^{(s)}]^4 + 6\sum_{s+l}(1-4a\delta)^{s+l}[v^{(s)}v^{(l)}]^2\right\}$$

$$= 3\sigma^4 a^4\left\{\sum_{s=0}^{\infty}(1-4a\delta)^{4s} + 2\sum_{s+l}(1-4a\delta)^{s+l}\right\} = \frac{3\sigma^4 a^4}{[1-(1-4a\delta)^2]^2} = 3D_x^2.$$

Thus, $M(x^4) = m_x^4 + 6m_x^2 D_x + 3D_x^2$.

Subtracting the expression (2.8) for $(M(x^2))^2$, we obtain

$$D_y = 4m_x^2 D_x + 2D_x^2. \tag{2.9}$$

Figure 2 shows the possible graphs of $\frac{m_y}{\sigma^2/\gamma^2}\left(\frac{a\gamma}{4}\right)$ (solid lines) and $\frac{1}{8\gamma^2}\frac{D_y}{\sigma^2/\gamma^2}\left(\frac{a\gamma}{4}\right)$ (dotted lines) for the case $\frac{\gamma}{\sqrt{\sigma}} = 0.2$. (the value taken from [2] for different values of a_0 ($a_0 = 0.5\gamma$, $a_0 = \gamma$, $a_0 = 2.5\gamma$). In constructing graphs of the dispersion it is assumed that $\gamma = 0.2$.

From the graphs it is clear that when a_0 increases the curves $m_y(a)$ become narrower, their minimal value decreases, and the minimum itself is shifted to the side of decreasing a. Clearly the minima $m_y(a)$ correspond to the

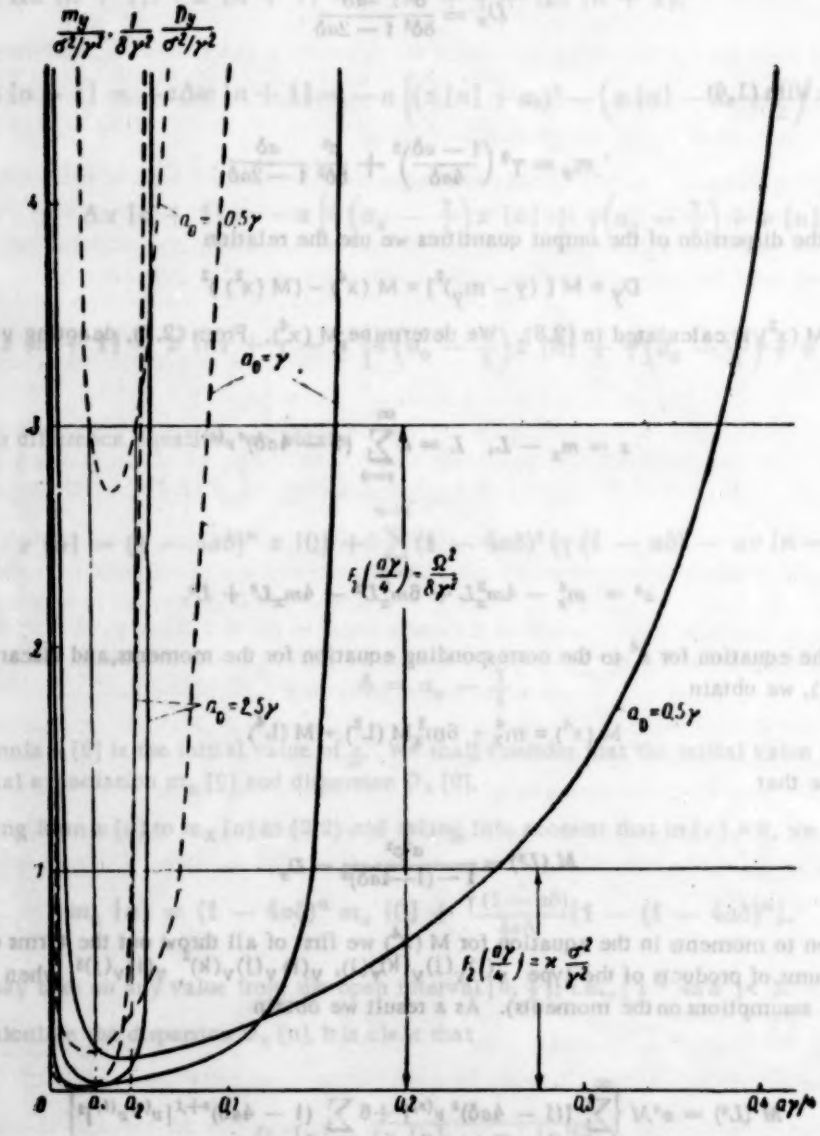


Fig. 2.

optimal (least m_y) values of the parameters $\underline{a}$ for the given a_0. In this, at the optimal point, the greatest a_0 guarantees the greatest step size with the least possible error (for large a_0 the influence of the disturbances is less). This explains the decrease in the minimal value of m_y with increasing a_0. However, when one deviates from the optimum, the large magnitude of the step may (although more seldom the less a_0 is) significantly deviate the system from it minimum. This explains the fact that with increaseing a_0 the curves $m_y(a)$ contract.

The dispersion curves change analogously. They, as well as the curves m_y, have asymptotes at the points $a\delta = 0$, $a\delta = \frac{1}{2}$ (with increasing a_0 the asymptotes are displaced in the direction of decreasing a). The minimum $D_y(a)$ disappears with increasing a_0 (the optimal choice of $\underline{a}$ for a given a_0 corresponds to the least possible error, if a_0 is large, and, consequently, the least spread relative to the mean value, i.e., the least dispersion).

If one requires that the condition

$$m_y < \varkappa \tag{2.10}$$

be satisfied, and also that

$$D_y < \Omega^2,$$

then these determine the region of possible values of the parameters $\underline{a}$, a_0. It is clear that the required control for a given a_0 may be secured only in the case when the minima of the curves m_y and D_y lie below the corresponding given values $\varkappa$ and Ω^2.

As examples, the lines $F_1\left(\frac{a\gamma}{4}\right) = \frac{\Omega^2}{8\gamma^2}$, $F_2\left(\frac{a\gamma}{4}\right) = \frac{\varkappa}{\sigma^2/\gamma^2}$ are constructed in Fig. 2. For $a_0 = 0.5\gamma$ the region of admissible $\underline{a}$ is included between the values a_1 and a_2, which are roots of the equation $D(\gamma)|_{a_0=0.5\gamma} = \Omega^2$. One may analogously determine the region of admissible $\underline{a}$ for $a_0 = \gamma$ and $a_0 = 2.5\gamma$. It is clear that the region of admissible $\underline{a}$, largest for $a_0 = \gamma$, shrinks as a_0 either increases or decreases.

Thus, when limitations on the steady values of the mathematical expectation and dispersion of the output quantities are prescribed, one may, knowing γ and σ, determine the possible values of the operating parameters.

We attempt to compare the systems with differing $\underline{a}$, a_0 in relation to their rapidity of action. We consider the magnitude of $m_y[n]$ [see (1.6), (2.4), and (2.5)]:

$$m_y[n] = (1-4a\delta)^{2n}[D_x[0] - D_x + (m_x[0] - m_x)^2] + (1-4a\delta)^n 2m_x(m_x[0] - m_x) + m_y. \tag{2.11}$$

It is evident that at the various moments of time $\underline{n}$ and depending on the values of the parameters a, δ and the initial conditions ($m_x[0]$, $D_x[0]$), the quantity $m_y[n]$ will be greater or less than m_y. Consequently, the regime of stability may be either aperiodic or oscillatory. Therefore, as quantities characterizing the stabilizing rate of the process, we consider the limit ($n \to \infty$) of the sum of squares of the deviation $m_y[n]$ from the mean values (use of a similar estimate was proposed by A. A. Feld'baum):

$$I = \lim_{n\to\infty} \sum_{k=0}^{n} (m_y[k] - m_y)^2. \tag{2.12}$$

We shall calculate the most rapidly acting system with the least I_0.

Setting expression (2.11) into (2.12), using the condition $0 < a\delta < \frac{1}{2}$, and denoting $a\delta = \theta$, we obtain

$$I = \frac{\left[\left(m_x[0] - \frac{\gamma(1-\theta)}{4\theta}\right)^2 + D_x[0] - \frac{\sigma^2}{8\delta^2}\frac{\theta}{1-2\theta}\right]^2}{1-(1-4\theta)^4} + \frac{4\gamma(1-\theta)}{4\theta}\frac{\left(m_x[0] - \frac{\gamma(1-\theta)}{4\theta}\right)\left[(m_x[0]-m_x)^2 + D_x[0] - \frac{\sigma^2}{8\delta^2}\frac{\theta}{1-2\theta}\right]}{1-(1-4\theta)^3} + 4\left(\frac{\gamma(1-\theta)}{4\theta}\right)^2\frac{(m_x[0]-m_x)^2}{1-(1-4\theta)^2}. \tag{2.13}$$

In the calculations it is assumed that the initial conditions are not random, i.e., $x_0 = s\gamma$. For large $\underline{s}$ one may suppose that

$$\frac{I_{ap}}{\gamma^4} = \frac{s^4}{1-(1-4\theta)^4} + \frac{4s^3}{1-(1-4\theta)^3}\frac{1-\theta}{4\theta} + \frac{4s^2}{1-(1-4\theta)^2}\left(\frac{1-\theta}{4\theta}\right)^2.$$

The graph of I_{ap}/γ^4 is given in Fig. 3 ($s = 100$). For a fairly wide range in θ ($0.10 < \theta < 0.40$) the quantity I changes by something like 1% from its minimal value.

The calculation by the exact formula was carried out for $a_0 = \gamma$, $a_0 = 2.5\gamma$, $a_0 = 5\gamma$. The graphs are given in Fig. 4. It is clear that they are practically identical with the graph of Fig. 3. The difference is only observed at the ends of the interval $[0, \frac{1}{2}]$ for $a_0 = 2.5\gamma$ and $a_0 = 5\gamma$. The graph for $a_0 = \gamma$ for large θ is somewhat lower

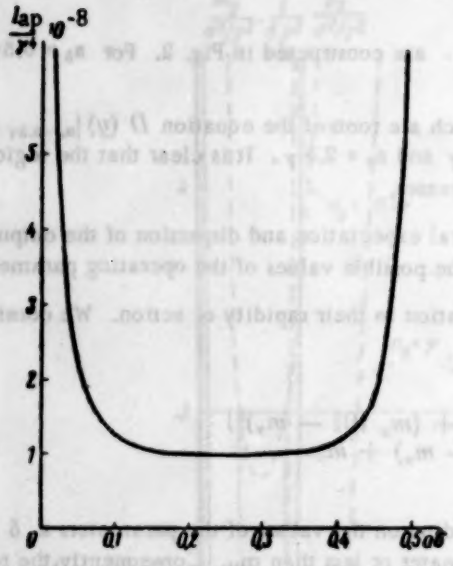


Fig. 3.

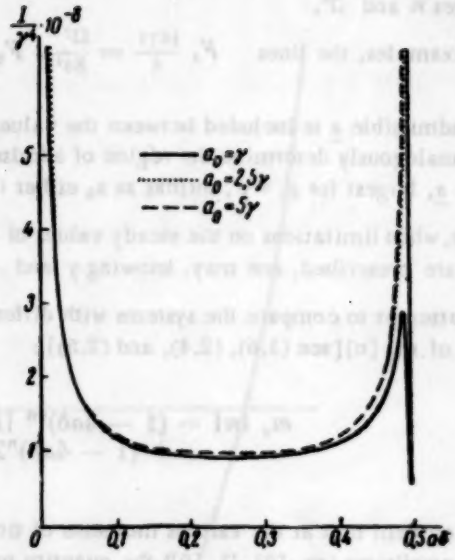


Fig. 4.

than the others and at the end of the interval has an additional narrow deep minimum. Clearly for moderate $\underline{s}$ and differing δ the shape of the curves may differ significantly from that in Fig. 4.

3. Case of m Independent Inputs

Proceeding from the results of section 1, we write the following equations for the input coordinate:

$$x_i[n+1] = x_i[n] + \gamma - a\left\{\sum_{j=1}^{m}\left(x_j[n] + \frac{2i-2}{2m}\gamma + a_0\delta_{ij}\right)^2 - \sum_{j=1}^{m}\left(x_j[n] + \frac{2i-1}{2m}\gamma - a_0\delta_{ij}\right)^2 + v_i[n]\right\} \quad (i = 1, 2, \ldots, m),$$

where

$$\delta_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j, \end{cases}$$

or

$$x_i[n+1] = x_i[n]\left(1 + \frac{a\gamma}{m} - 4aa_0\right) + \frac{a\gamma}{m}\sum_{j\neq i} x_j[n] + \gamma\left(1 - \frac{4i-3}{m}a\delta\right) - av_i[n] \quad (i = 1, 2, \ldots, m),$$

where

$$\delta = a_0 - \gamma/4. \tag{3.1}$$

We introduce the notation:

$$\alpha = 1 - 4aa_0 + \frac{a\gamma}{m}, \quad \beta = \frac{a\gamma}{m}, \quad \varkappa_i[n] = \gamma\left(1 - \frac{4i-3}{m}a\delta\right) - av_i[n]. \tag{3.2}$$

In these terms, in order to determine x_i [n] we obtain a system of difference equations

$$x_i[n+1] = \alpha x_i[n] + \beta \sum_{j \neq i} x_j[n] + \varkappa_i[n] \quad (i = 1, 2, \ldots, m). \tag{3.3}$$

We find the mathematical expectation $m_i[n]$ of the magnitude of the input coordinate of the system $x_i[n]$. Employing the fact that $M(v) = 0$, we obtain the following system of difference equations:

$$m_i[n+1] = \lambda_i + \sum_{j=1}^{m} \alpha_{ij} m_j[n] \quad (i = 1, 2, \ldots, m),$$

where

$$\lambda_i = \gamma\left(1 - \frac{4i-3}{m} a\delta\right), \quad \alpha_{ii} = \alpha, \quad \alpha_{ij}|_{i \neq j} = \beta \tag{3.4}$$

or the vector equation

$$(\hat{I} + \hat{A}E_{n+1}^{-1})\,\mathbf{m}[n+1] = \boldsymbol{\lambda}, \tag{3.5}$$

where $\hat{I}$ is the unit matrix,

$$-\hat{A} = \|\alpha_{ij}\| = \begin{pmatrix} \alpha & \beta & \ldots & \beta \\ \beta & \alpha & \ldots & \beta \\ \ldots & \ldots & \ldots & \ldots \\ \beta & \beta & \ldots & \alpha \end{pmatrix}$$ is a matrix with m rows and m columns,

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_m \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix},$$

E is the operator of displacement in unit time (E^{-1} is backward displacement in unit time).

We obtain the following solution for the current time n:

$$\mathbf{m}[n] = (-\hat{A})^n \mathbf{m}[0] + \sum_{s=0}^{n-1} (-\hat{A})^s \boldsymbol{\lambda}, \tag{3.6}$$

in which m [0] corresponds to x [0], where x [0] is the initial value vector of the input variables.

For the stable regime ($n \to \infty$) [3]

$$\mathbf{m} = \sum_{s=0}^{\infty} (-\hat{A})^s \boldsymbol{\lambda}. \tag{3.7}$$

We represent $-\hat{A}$ in the form

$$-\hat{A} = (\alpha - \beta)\hat{I} + \beta\hat{E},$$

where

$$\hat{E} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & \ldots & 1 \end{pmatrix}.$$

Then in correspondence with (3.7)

$$\mathbf{m} = \sum_{s=0}^{\infty} [(\alpha - \beta)\hat{I} + \beta\hat{E}]^s \lambda = \sum_{s=0}^{\infty} \left\{ \sum_{k=0}^{s} C_s^k [(\alpha - \beta)\hat{I}]^k (\beta\hat{E})^{s-k} \right\} \lambda.$$

It is easy to see that $\hat{E}^{(r)} = m^{r-1}\hat{E}$, hence

$$\mathbf{m} = \sum_{s=0}^{\infty} \left\{ \sum_{k=0}^{s-1} C_s^k (\alpha - \beta)^k \frac{(\beta m)^{s-k}}{m} \hat{E} + (\alpha - \beta)^s \hat{I} \right\} \lambda$$

$$= \sum_{s=0}^{\infty} \left\{ \left[\sum_{k=0}^{s} C_s^k (\alpha - \beta)^k (\beta m)^{s-k} - (\alpha - \beta)^s \right] \frac{\hat{E}}{m} + (\alpha - \beta)^s \hat{I} \right\} \lambda$$

$$= \sum_{s=0}^{\infty} \frac{(\alpha - \beta + \beta m)^s - (\alpha - \beta)^s}{m} \hat{E}\lambda + \sum_{s=0}^{\infty} (\alpha - \beta)^s \hat{I}\lambda.$$

The solution is finite in the case when

$$|\alpha - \beta + \beta m| < 1, \quad |\alpha - \beta| < 1 \tag{3.8}$$

or

$$0 < a\delta < \tfrac{1}{2}, \quad 0 < aa_0 < \tfrac{1}{2}. \tag{3.9}$$

Supposing that condition (3.8) is fulfilled; it follows that

$$m_i = \frac{\beta \sum_{j=1}^{m} \lambda_j}{(1 - \alpha + \beta - \beta m)(1 - \alpha + \beta)} + \frac{\lambda_i}{1 - \alpha + \beta}.$$

This same expression is obtained for the quantity J_i in [2] if one solves the system (44) in [2] under the assumption that $b_{jk} = \delta_{jk}$.

After certain transformations we may write

$$m_i = \frac{\gamma}{4aa_0} \left\{ \left[\frac{\gamma}{4\delta} \left(1 + \frac{a\delta(1 - 2m)}{m} \right) + 1 + \frac{3a\delta}{m} \right] - i\frac{4a\delta}{m} \right\}. \tag{3.10}$$

We consider the magnitude of the dispersion of the i-th input coordinate: $D_i = M[(x_i - m_i)^2]$. Solving equation (3.3), we obtain

$$\lim_{n \to \infty} x_i[n] - m_i = \sum_{s=0}^{\infty} \left\{ -\frac{(1 - 4a\delta)^s - (1 - 4aa_0)^s}{m} a \sum_{l=1}^{m} v_l^{(s)} - (1 - 4aa_0)^s a v_i^{(s)} \right\},$$

where $\varkappa_l^{(s)}$ and $v_l^{(s)}$, respectively, denote $\varkappa_l[n - s]$ and $v_l[n - s]$. Squaring the latter expression and integrating with density $\rho(v)$ with respect to v ($v = \{v_1[0], \ldots, v_m[0], v_1[1], \ldots, v_m[n]\}$), we obtain the following expression for the dispersion:

$$D_i = D_x = \frac{a^2\sigma^2}{8m} \left[\frac{1}{a\delta(1 - 2a\delta)} + \frac{m - 1}{aa_0(1 - 2aa_0)} \right]. \tag{3.11}$$

From (3.11) it is clear that the dispersions for all input coordinates are identical. The quantity m_y is determined from (1.6) with account taken of (3.10) and (3.11).

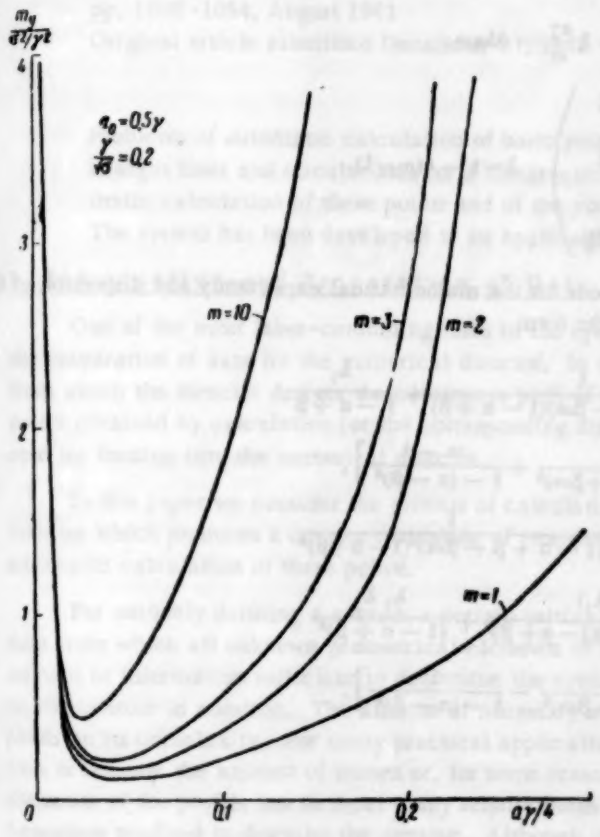


Fig. 5

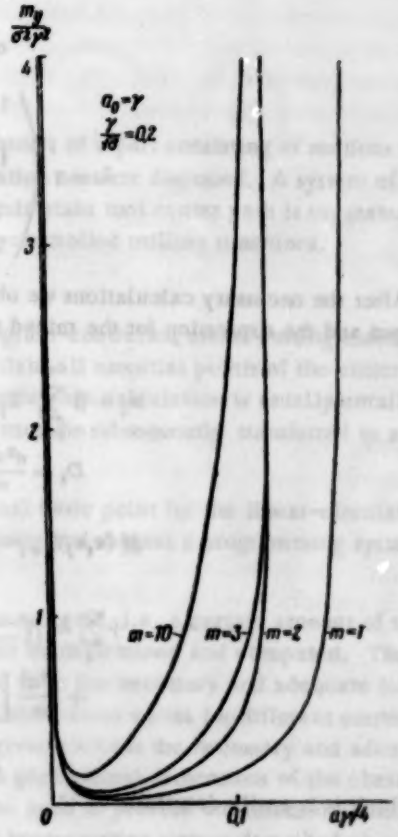


Fig. 6

In Figs. 5, 6 we have graphs of $m_y/\sigma^2/\gamma^2$ depending on $a\gamma/4$ for m = 1,2,3,10 for different a_0. From the graphs it is evident that when the number of inputs is increased, the steady value of the mathematical expectancy of the output quantity rises (for a fixed a).

The minimal values of m_y increase with increasing m and the minimum itself is shifted to the left along the a axis.

In principle there is no difficulty posed in the consideration of the dispersion of the output quantity for a system with m independent inputs, nor in the estimation of the rapidity of action of the latter in relation to the criterion introduced in section 2. However, it involves very detailed calculations.

The method of calculating m_y in this section may easily be generalized to the case of a system with m independent inputs, whose input and output signals are connected by the relation

$$y[n] = \sum_{i=1}^{m} x_i^2[n] + b\sum_{i\neq j} x_i[n]x_j[n].$$

The vector equation for $x_i[n]$ has the form

$$x[n+1] = \hat{A}x[n] + \lambda - av[n],$$

where

$$\hat{A} = \|a_{ij}\|, \quad a_{ii} = \alpha = 1 - 4aa_0 + \xi\frac{a\gamma}{m},$$

$$a_{ij}|_{i\neq j} = \beta = \xi\frac{a\gamma}{m} - 4baa_0,$$

$$\lambda = \gamma\begin{pmatrix} 1-\frac{4\cdot 1-3}{m}a\delta\xi \\ 1-\frac{4\cdot 2-3}{m}a\delta\xi \\ \cdots\cdots\cdots \\ 1-\frac{4\cdot m-3}{m}a\delta\xi \end{pmatrix}, \quad \xi = 1 - b(m-1).$$

After the necessary calculations we obtain expressions for the mathematical expectancy and dispersion of the i-th input and the expression for the mixed moment in the form

$$m_i = \beta\sum_{j=1}^{m}\lambda_j\frac{1}{(1-\alpha+\beta-\beta m)(1-\alpha+\beta)} + \frac{\lambda_i}{1-\alpha+\beta},$$

$$D_i = \frac{a^2\sigma^2}{m}\left[\frac{1}{1-(\alpha-\beta+\beta m)^2} + \frac{m-1}{1-(\alpha-\beta)^2}\right],$$

$$M(x_ix_j)|_{i\neq j} = \beta^2\sum_{k,l}\lambda_k\lambda_l\frac{1}{(1-\alpha+\beta-\beta m)^2(1-\alpha+\beta)^2}$$

$$+\sum_l\lambda_l\frac{\beta(\lambda_i+\lambda_j)}{(1-\alpha+\beta-\beta m)(1-\alpha+\beta)^2} + \frac{\lambda_i\lambda_j}{(1-\alpha+\beta)^2}$$

$$+\frac{a^2\sigma^2}{m}\left[\frac{1}{1-(\alpha-\beta+\beta m)^2} - \frac{1}{1-(\alpha-\beta)^2}\right],$$

from which it is easily determined that

$$m_y = \sum_{i=1}^{m}(m_i^2 + D_i) + b\sum_{i\neq j}M(x_ix_j).$$

Clearly the analysis of the dependence of m_y on the operating parameters will be fairly complicated.

Thus, in the case of a system with m independent inputs, as in the case $m = 1$, operating formulas and graphs are obtained, permitting one to determine the possible sets of parameters a, a_0, if the maximal admissible values of m_y are given.

In conclusion the author thanks A. A. Fel'dbaum for his attention to this paper and for useful comments, and also A. I. Teiman for discussions.

LITERATURE CITED

1. Pugachev, V. S. Theory of Random Functions and its Application to Automatic Control Problems. Fizmatgiz, 1960.
2. Fel'dbaum, A. A. Statistical theory of gradient systems of automatic optimization with square object characteristics. Avtomatika i Telemekhanika, Vol. XXI, No. 2, 1960.
3. Bartlett, M. S. Introduction to the Theory of Random Processes [Russian translation], 1958.

A SYSTEM FOR AUTOMATIC PROGRAMMING FOR MACHINE TOOLS

B. G. Tamm (Tallinn)

Translated from Avtomatika i Telemekhanika Vol. 22, No. 8,
pp. 1038-1054, August 1961
Original article submitted December 27, 1960

Problems of automatic calculation of basic points on the contours of a part consisting of sections of straight lines and circular arcs for a linear-circular interpolation unit are discussed. A system of automatic calculation of these points and of the points on the equidistant tool center path is suggested. The system has been developed to be applicable to program-controlled milling machines.

1. Introduction and Preparation of Data

One of the most labor-consuming tasks in the operation of program-controlled, metal cutting machine tools is the preparation of data for the numerical director. In order to calculate all essential points of the cutter center path, from which the director derives the continuous path of the tool, considerable calculation is usually entailed. All points obtained by calculation (or the corresponding displacements) must be subsequently transferred to a program card for feeding into the numerical director.

In this paper we consider the process of calculation of individual basic point for the linear-circular numerical director which produces a contour consisting of straight lines and circles and suggest a programming system for the automatic calculation of these points.

For uniquely defining a contour a certain initial information is required, i.e., a certain amount of numerical data from which all unknown geometrical elements of the profile can be determined and computed. The minimum amount of information sufficient to determine the contour is referred to as the necessary and adequate information on the contour in question. The amount of necessary and adequate information varies for different contours and depends on its complexity. For many practical applications the data given exceeds the necessary and adequate amount. This is because the amount of known or, for some reason, calculated geometrical dimensions of the characteristic elements of the profile has to meet many requirements other than the need to provide the minimum amount of information required to describe the contour. Although the automatic programming system described above can always compute a contour using the necessary and adequate data, an excess of original information never interferes with the calculation and always makes the work of determining the unknown elements of the profile easier for the programmer. It should also be noted that it is always preferable to have for an automatic programming system data which are uniformly distributed over the entire contour. If this is not the case it may happen that some elements can be determined only indirectly, which in subsequent calculations can produce an accumulated error.

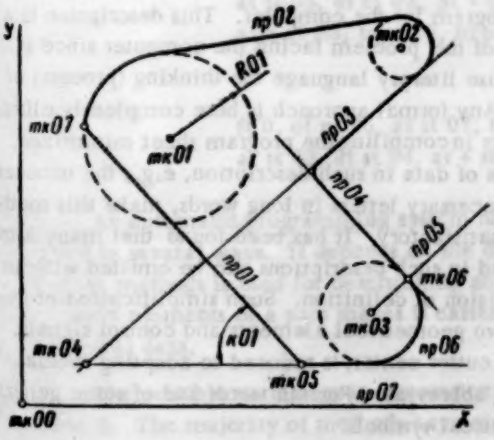


Fig. 1.

Figure 1 shows a relatively simple part consisting of short straight lines and circular arcs. The coordinates of five points, one radius, and the inclination of one straight line are given and determine the whole contour. For determining the path of the cutter center during the machining it is first necessary to calculate the coordinates of seven points of the profile. The problem involves the calculation of equations of seven straight lines, the radii of two circles, and the solution of several systems of linear equations. Almost the same amount of work is involved in the subsequent calculation of the equidistant curve to the given contour and in entering the results obtained on the program card.

For making the preparation of data automatic and to relieve the operator from this work, in both the solution and the setting up of algorithms, automatic methods must be adopted and the translation of the data into the "machine language" made simpler. This can be successfully done by using general-purpose digital computers equipped with automatic programming units.

In this case all unknown geometrical elements can be easily determined indrectly from known elements. On the blueprint of the part illustrated in Fig. 1 the known (given as figures) elements are given in bold print. The straight line 01, which is not known, is defined by point 05 on this line and the slope to 01; point 07 is determined as the point of intersection of straight line 01 and circle 01 (the point with a larger ordinate is selected from the two points of intersection); straight line 02 is defined as the external tangent to circles 01 and 02; circle 01 is determined by the coordinates of its center (point 01) and radius 01 while circle 02 is given by the coordinates of its center (point 02) and one tangent (straight line 03). Line 03 can, in turn, be determined as the straight line perpendicular to line 01 (which has been calculated earlier) passing through the known point 04. All other elements needed for determing the contour are found in a similar way.

Let us now assume that all elements have been calculated. Next, it is necessary to establish the sequence of required motions of the cutter center with respect to these elements. The path shown in Fig. 1 is machined by milling so that the path of the cutter center must be described as follows.

The cutter motion begins at point 00 in the origin of the coordinate system and proceeds along the straight line 07 to line 01. The motion continues along line 01 but to the left from it at a distance equal to the cutter radius. So it moves to point 07 and then along the circular arc 01 in clockwise direction until it meets line 02; it then moves along this line to meet the circular arc 02, etc., until it returns to point 00.

In handling this problem by conventional methods it is first necessary to calculate all above-mentioned indirectly determined points on the contour and on the cutter path. This is usually done on a keyboard or a general-purpose computer. Thereupon all results must be entered on the program card from which they are transferred by punching onto a tape and fed into a general-purpose computer, which is then in position to produce a program for the interpolation unit. The procedure involves a considerable amount of work as it requires the processing of systems of equations and repeated rewriting of a considerable number of figures. In addition, the rewriting is a potential source of errors.

The suggested automatic programming system incorporates a complicated component program which enables the required algorithms to be set up. Their subsequent solutions are found by means of an ordinary general-purpose computer. The implementation of this system calls for a standardization of calculations of all typical problems; the logics of the system must be able to decipher all possible combinations of these problems and to set up the corresponding algorithms.

TABLE 1

Abbreviation	Meaning of the abbreviation
tk (Eng.: pt)	tochka (full stop)
pr (st)	pryamaya (straight line)
kr (ci)	okruzhnost' (circle)
stop	stop, ostanovka (stop, stoppage)
∥	parallel'no (parallel)
⊥	perpendikulyarno (perpendicular)
b (m)	bol'she (more)
m (l)	men'she (less)
ot (fr)	ot (from)
do (to)	do (to)
po (al)	po (along)
R	radius (radius)
fr (cu)	freza (milling cutter)
fr+ (cu+)	freza sprava (cutter to the right)
fr- (cu-)	freza sleva (cutter to the left)
k (k)	tangens ugla naklona pryamoi linii (tangents of the angle of slope)
+kr (+ci); -kr (-ci)	dvizhenie po okruzhnosti po chasovoi strelke (clockwise circular motion)

One of the important points in developing the automatic programming system (APS) was to decide how to feed the original information into the computer. This information must describe all unknown geometrical elements in terms of known elements and also control the entire calculation process. We described above for the part illustrated in Fig. 1 how some of the unknown elements are expressed in terms of known elements. The description of the path of the cutter center, which has also been partially described, is in fact the control program for the computer. This description is a natural expression of the problem facing the computer since it expresses in a precise literary language the thinking (process) of the programmer. Any formal approach is here completely eliminated and the errors in compiling the program sheet minimized. However, an excess of data in such description, e.g., the unnecessary words and unnecessary letters in long words, make this method of feeding data unsatisfactory. It has been found that many letters and words contained in such descriptions can be omitted without damaging the precision of definition. Such simplification of description of unknown geometrical elements and control signals, and of the path of cutter center, is reduced to adopting special terms consisting of abbreviated Russian words and of some generally accepted mathematical symbols.

The terms obtained by abbreviating Russian words used in this paper are given in Table 1. In future these expressions may

be considered as words of a special programming language, while the sentences compiled of these words could be referred to as the language of the automatic programming system.

In making this suggestion the words of the language are used in the same way as usual in forming literary sentences. The only substantial difference is that in the programming language the same terms must always be expressed by the same words; their meaning is strictly defined and standardized. The sentences written in APS language which determine the geometrical elements of the part are referred to as "designations". All designations are divided into two groups: "elementary designations" and "complex designations". The elementary designations are designations used for entering the numerical data into a drawing. In our system these designations include points with their coordinates, radii of circles, and the slopes of straight lines expressed in ordinary language. Complex designations adopted for APS make possible the description of all geometrical elements of a path given indirectly, i.e., elements which can be calculated from the original numerical data by means of the APS.

All sentences needed for defining the contour of a part are given in form of elementary and complex designations in the first two sections of Table 2.

TABLE 2

Elementary designations

X pt 00 = 00	Y pt 00 = 00	R 01 = 15
X pt 01 = 26	Y pt 01 = 47	
X pt 02 = 66	Y pt 02 = 63	
X pt 03 = 60	Y pt 03 = 16	
X pt 04 = 11	Y pt 04 = 07,5	
X pt 05 = 48		K 01 = − 0,9

Complex designations

pt 00 = X00/00	st 01 = − pt 05/01	ci 01 = pt 01/R 01
pt 01 = X01/Y01	st 02 = + ci 01/+ ci 02	ci 02 = pt 02/st 03
pt 02 = X02/Y02	st 03 = ⊥ st 01/pt 04	ci 03 = pt 03/st 06
pt 03 = X03/ Y03	st 04 = ∥ st 01/pt 06	
pt 04 = X04/Y04	st 05 = ⊥ st 01/pt 03	
pt 05 = X05/Y04	st 06 = Y04	
pt 06 = M X ci 03/st 05	st 07 = Y00	
pt 07 = M Y ci 01/st 01		

Control program

Variant I

ft 0, of pt 00, at st 07, to st 01, ft −, at st 01, to pt 07, at + ci 01, to st 02, at st 03, to ci 02, at + ci 02, to st 03, at st 03, to st 04, at st 0.5, at ci 03, at + ci 03, to st 06, to pt 05, to pt 00 stop.

Variant II

ft 0, of pt 00, at st 07, ft −, at st 01, to pt 07, at + ci 01, at st 02, at + st 02, at st 03, at st 04, at + st 03, at st 06, to pt 04, to pt 00, stop.

An automatic programming system has one important feature: The same geometrical element can be determined in several ways. It depends on the data available and also on the choice and taste of the programmer which of these methods is used for determining an element. The presence of several complex designations for determining the same elements of a part makes it easier for the programmer to determine the entire part using a small number of numerical data.

The possible variations in denotation of points, straight lines, and circles are given for the described system in Table 3. The majority of methods have several variants. For example, in the case of an intersection of a straight line with a circle or of two circles, the solution of relevant equations yields two roots. For finding the required root the command signal contains as an additional element one of the coordinates of the point in question with respect to

TABLE 3

A. Points

1. pt 01 = X 01/Y 01	By Cartesian coordinates
2. pt 01 = st 01/st 02	As intersection of two straight lines
3. pt 0 = M X ci 01/ st 01	As intersection of a straight line and a circle
4. pt 01 = c X ci 01/st 01	
5. pt 01 = M Y ci 01/st 01	
6. pt 01 = c Y ci 01/st 01	
7. pt 01 = M X ci 01/ci 02	As intersection of two circles
8. pt 01 = c X ci 01/ci 02	
9. pt 01 = M Y ci 01/ci 02	
10. pt 01 = c Y ci 01/ci 02	

B. Straight lines

1. st 01 = ± pt 01/K 01	By one point and the slope
2. st 01 = pt 01/pt 02	By two points
3. st 01 = ± ci 01/± ci 02	As a tangent on 2 circles
4. st 01 = X 01	By its constant abscissa
5. st 01 = Y 01	By its constant ordinate
6. st 01 = M X st 02/R 01	As a straight line parallel to the known straight line shifted through a definite distance with respect to the latter
7. st 01 = c X st 02/R 01	
8. st 01 = M Y st 02/R 01	
9. st 01 = c Y st 02/R 01	
10. st 01 = ∥ st 02/pt 01	As former but passing through a given point outside this straight line
11. ⊥ st 01 = 02/pt 01	As the perpendicular through the known straight line passing through the known point

C. Circles

1. ci 01 = pt 01/R 01	By the coordinates of the center and radius
2. ci 01 = pt 01/st 01	By the coordinates of the center and one tangent
3. ci 01 = pt 01/pt 02	By the coordinates of the center and one point of the circle
4. ci 01 = pt 01/pt 02/pt 03	By three points on the circle
5. ci 01 = M X st 01/ M X st 02/R 01	By two tangents to the circle and the radius (a total of 16 different command signals)
6. ci 01 = c X st 01/M X st 02/R 01	
7. ci 01 = M Y st 01/M X st 02/R 01	
8. ci 01 = c Y st 01/M X st 02/R 01	
9.	
. . . . .	
. . . . .	
20. ci 01 = c Y st 01/c Y st 02/R 01	

the same coordinate of the second point. In the designation of sec. 3 division B in Table 3 (abr., B_3) the straight line is given as a tangent to its two circles; the signs in front of "ci" indicate the presence of an external or internal tangent. The number of methods given to denote three main geometrical elements is, of course, not final and is likely to increase as the APS develops.

It is evident that the description of the path of the tool's center contains all its required motions in the desired sequence. In order to make this description a control program of the APS it is necessary to record it in such a manner that its decoding by the computer produces the action of the APS needed for the calculation of the cutter center path and for preparing the program for the numerical director. This can again be made by using the terms of the programming language. A detailed description of the cutter path written in the programming language is called the control

program. The control program (CP) for the machining of the part illustrated in Fig. 1 is shown in the third section of Table 2. The control program together with a group of designations (which incorporates the elementary as well as complex designation) gives what is called the language program of the path.

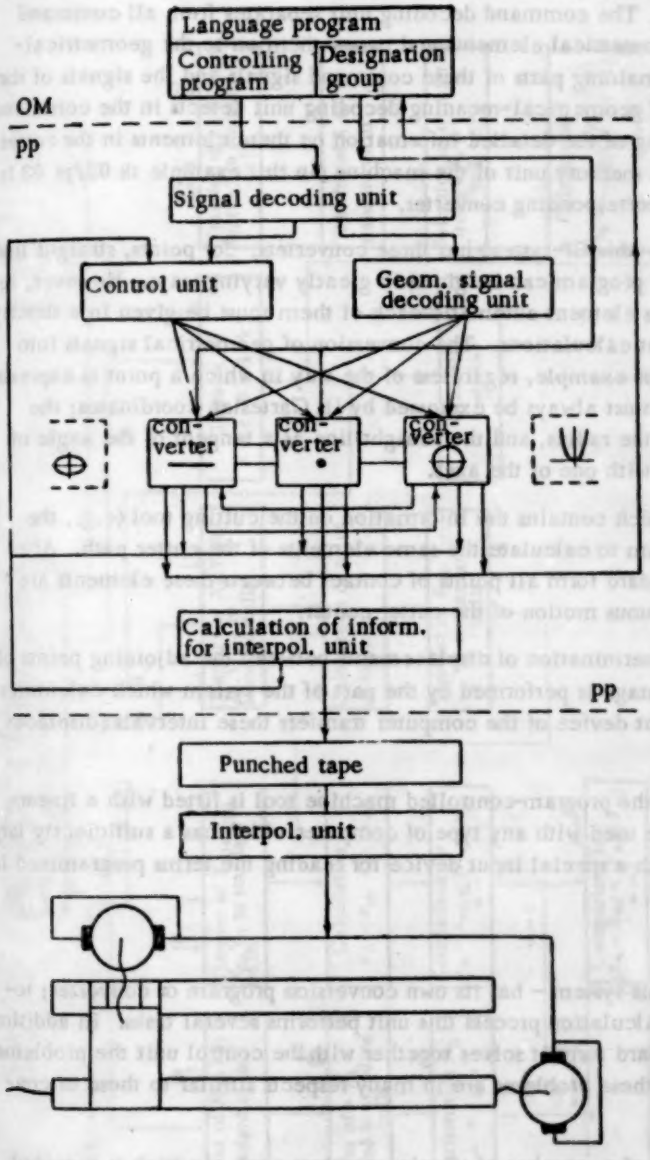


Fig. 2.

The setting up of the control program is in principle similar to determining designations. In contrast to notations which represent abridged sentences in program language ,which describe the geometrical elements of the path, the individual command signals of the CP separated by commas represent the corresponding sentences which describe the motion of the cutter center in the abbreviated form (see Table 2 and the description of the motion of the cutter center on page 1039). The separation of individual command signals by a comma is used to facilitate the reading of the program and to avoid errors. The depressed "comma" key of the punching machine corresponds to the signal "recording". This ensures the recording of data contained between two commas into one memory cell of the computer.

Before feeding into the computer the language program is punched on a special machine. This machine has a keyboard similar to that of a typewriter and contains all characters and symbols necessary to reproduce the language program. The program is simultaneously recorded as a punched tape in form of combinations of eight or ten-number figures and typed on paper in APS language. The elementary data are transferred on the ten-channel (decimal-system) punched tape and so fed into the computer. The complex designations and CP are reproduced and fed into the machine on eight-channel system tape. Only the right-hand sides of designations are fed into the computer. Groups of information data consisting of elementary or complex designations are called "similar" if they characterize geometrical elements of the same kind. The decimal-system and 8-channel-system information is fed into the memory unit of the machine as solid groups and are marked by symbols derived from $X_n - 1$, where X_n is the number of the cell containing the first term of the n-th similar information group.

A group of denotations is introduced into the operational memory of APS and the signals of the control program are read gradually by the counting device as the calculation proceeds. The automatic programming system, which is described and discussed in the next two sections, is permanently in the continuous memory(or in the memory of constants) of the digital computer.

Before proceeding to examine in greater detail the individual parts of the APS we shall consider briefly the general layout of the system. A simplified block diagram of APS is shown in Fig. 2. The calculation process can begin after introducing a group of designation symbols into the operational memory unit. The signal of the control program proceeds after reading, by the input device, into the control unit of the automatic programming system. The control unit (CU) consists of program units which perform the general logical operation of the system, the automatic setting up of algorithms of calculations, and also the decoding of input information. In order to make the ex-

planation of the operation of the system easier the control units concerned with the decoding of command signals of CP are shown in Fig. 2 detached from the rest of the CU. These are the decoding unit of command signals and the decoding unit of the geometrical meaning of command signals.

The input unit passes the command signals of CP one by one into the decoding unit which divides all signals into two groups. The first group includes signals without any geometrical elements (e.g., fr) while the second group comprises all other command signals (e.g., po + kr 02). The command decoding unit separates from all command signals of the second group parts which represent the geometrical elements and passes them on to the geometrical-meaning decoding unit (in our example kr 02). The remaining parts of these command signals and the signals of the first group are passed on into the other parts of CU. The geometrical-meaning decoding unit detects in the command signal the geometrical elements and arranges the reading of the detailed information on these elements in the corresponding group of complex designations contained in the memory unit of the machine (in this example tk 02/pr 03 is counted as kr 02). This information is passed onto the corresponding converter.

Every geometrical element has its own converter; this CP system has three converters: for points, straight lines, and circles. The geometrical elements in the language program can be given in greatly varying ways. However, to make all mathematical calculations concerned with this element automatic each of them must be given in a strictly defined form which can form the basis for the subsequent calculations. The conversion of geometrical signals into such standard form takes place in the converter unit. For example, regardless of the way in which a point is expressed in the language program at the converter unit input, it must always be expressed by its Cartesian coordinates; the circle is expressed by the coordinates of the center and the radius, and the straight line as a tangent of the angle of slope and the coordinate of the point of its intersection with one of the axes.

Thereupon the signal given by the control unit which contains the information on the cutting tool (e.g., the radius of the cutter, cutter on the left) makes the program to calculate the same elements of the cutter path. After converting the mean elements of cutter path into a standard form all points of contact between these elements are calculated which are needed for determining the continuous motion of the cutter center.

The last stage of the calculation process is the determination of displacements between the adjoining points of the cutter path measured on the coordinate axes. This stage is performed by the part of the system which calculates the data for the interpolating unit. Thereupon the output device of the computer transfers these intervals (displacements) onto the perforated tape.

Our automatic programming system assumes that the program-controlled machine tool is fitted with a linear-circular interpolating unit. The APS can in principle be used with any type of computer which has a sufficiently large rapid-access or constant memory and which is fitted with a special input device for reading the terms programmed in the APS language.

2. Converter Unit

Every geometrical element – there are three in this system – has its own conversion program or converter; together they are called the converter unit. During the calculation process this unit performs several tasks. In addition to converting the given geometrical element to a standard form it solves together with the control unit the problems involved in determining the cutter center path, because these problems are in many respects similar to those of conversion.

The converting programs represent a combination of a number of simple operators each of which is provided for solving a small highly specialized mathematical or logical problem. According to the command signal received at the input of the decoding unit, the control unit automatically synthesizes these operators to a sequence required for solving the problem. This system of synthesis of small operators is more flexible than the system using standard subprograms for the component.

After reading the detailed information on the geometrical element the information is passed on into the decoding unit of the corresponding converter. Each converter has its own decoding unit whose task it is to fix the method of determining one or other of the geometrical elements. In the system being considered these decoding units determine by means of a minimum number of comparisons any of the 17 methods of describing an element. The further, more detailed, decoding of the command signal is effected in the converters during the calculation process.

The decoding units are based on the principle of the minimum number of comparisons because the capacity of the memory of the M-3 general-purpose computer which was used in our experiments was limited. A shortcoming of

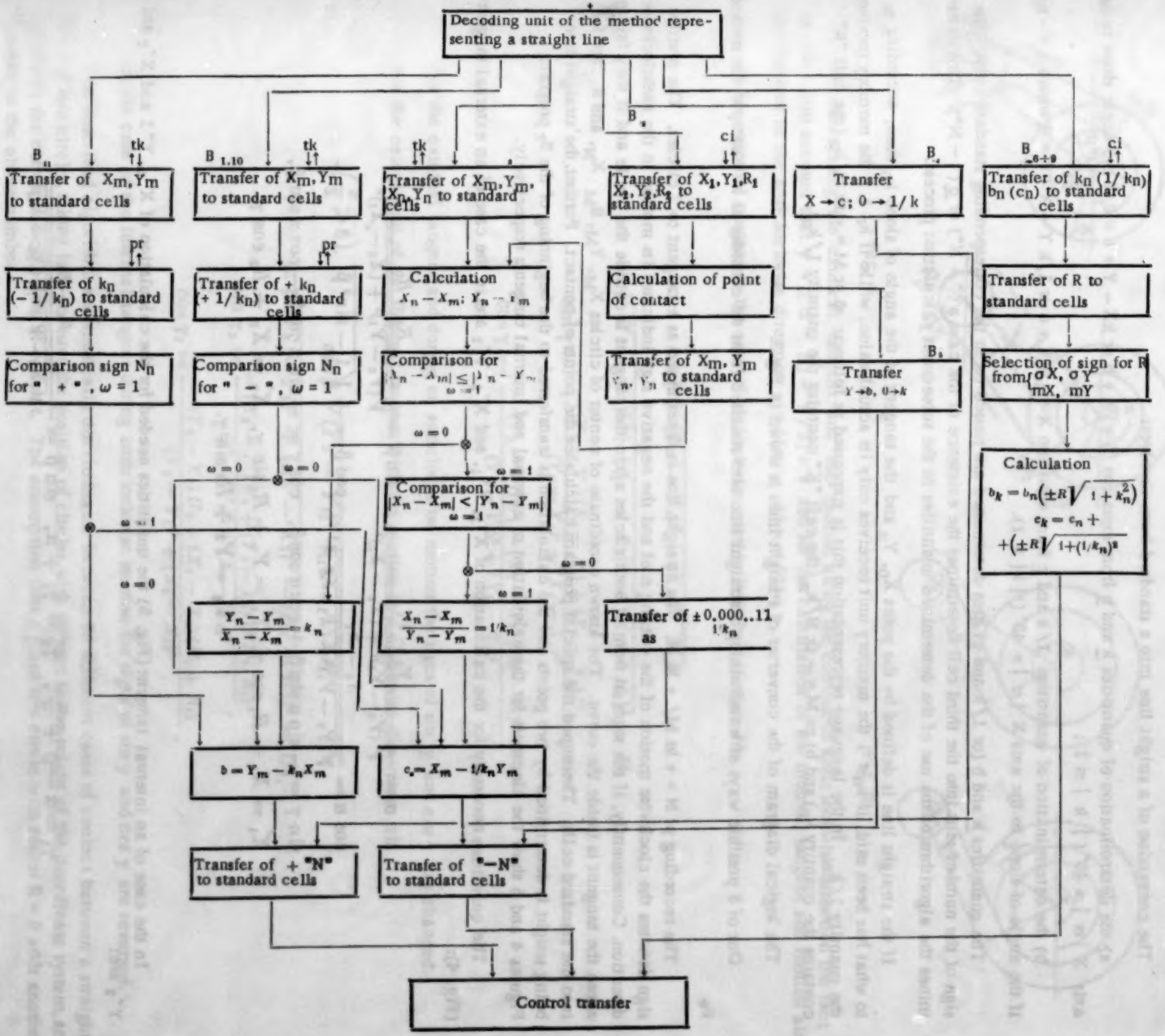

Decoding unit of the method representing a straight line
Transfer of Xm, Ym to standard cells
Transfer of Xm, Ym to standard cells
Transfer of Xm, Ym, Xn, Yn to standard cells
Transfer of X1, Y1, R1 X2, Y2, R2 to standard cells
Transfer X → c; 0 → 1/k
Transfer of kn (1/kn) bn (cn) to standard cells
Transfer of kn (− 1/kn) to standard cells
Transfer of + kn (+ 1/kn) to standard cells
Calculation
Calculation of point of contact
Transfer of R to standard cells
Comparison sign Nn for " + ", ω = 1
Comparison sign Nn for " − ", ω = 1
Comparison for
Transfer of Xm, Yn to standard cells
Transfer X → b, 0 → k
Selection of sign for R from {σX, σY, mX, mY
Comparison for |Xn − Xm| ≤ |Yn − Ym|, ω = 1
Calculation
Transfer of ±0.000..11 as 1/kn
Control transfer

Fig. 3.

these decoding units is that their operation is positively connected with the position of data in the discharge grid of the memory unit, which means that the adding of new designations would involve a complete reconstruction of the decoding unit circuit. If quick-acting machines with large memory capacity are used, more universal results are obtained with a coding unit which compares all double discharges of every designation. The sequence of comparisons should, of course, be fixed in such a manner that the more frequently encountered designations are determined first and the less likely designations are arranged at the end.

The conversion of a stright line into a standard form involves:

a) the determination of quantities $\underline{k}$ and $\underline{b}$ from equation $Y = kX + b$ or $kX - Y + b = 0$ if the angle slope to the axis $X \mid \alpha \mid \leq 45^\circ$ ($\mid k \mid \leq 1$);

b) the determination of quantities $1/k$ and $\underline{c}$ from equation $X = 1/kY + c$ or $1/k\, Y - X + c = 0$, where $c = -b/k$, if the angle of slope to the axis $X \mid \alpha \mid > 45^\circ$ ($\mid k \mid > 1$).

The quantities $\underline{k}$ and $\underline{b}$ (or $1/k$ and $\underline{c}$) thus determined are passed on to the corresponding standard cells. The sign of the number sent into the third cell determines the existence of the case $\underline{a}$ (" + N") or $\underline{b}$ (" − N"). This determines the algorithm of the use of the determined quantities in the subsequent calculation process.

If the straight line is defined by the point X_n, Y_n and the tangent of the angle of slope k_n then, according to to what has been said, if $k_n < 1$ the memory unit receives only its actual value, while if $k_n > 1$ the memory receives the quantity $1/k_n$. In the language representation this is expressed as follows: a) pr M = + tk N/k_n (the cell "K" contains the quantity k_n) and b) pr M = − tk N/k_n (the cell "K" contains the quantity $1/k_n$).

The logical diagram of the converter of straight lines is shown in Fig 3.

Out of 8 possible ways of transforming a straight line into standard form we consider as an example the method B_2.

The recording pr N = + kr M/ + kr N. The straight line is determined as a tangent to 2 circles. The positive sign denotes the clockwise motion of the cutting tool and the negative sign indicates its motion in the anticlockwise direction. Consequently, if the signs in front of both circles agree the tangent is outside the circle and if they fail to agree the tangent is inside the curve. The known coordinate of center of circles X_M, Y_M, R_M X_N, and R_N are fed into the standard cells. Thereupon the special program calculates the points of contact. Further, the straight line being sought is determined by two points and the calculation is transferred to the beginning of the B_2 program. Figures 4 and 5 show the elements for the calculation of external and internal tangents respectively.

The quantities necessary for the calculation of X'_1, Y'_1, and X'_2, Y'_2 are, in the case of an external tangent (Fig. 4):

$$\sin\alpha = \frac{Y_2 - Y_1}{\sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}}, \quad \sin\beta = \frac{R_2 - R_1}{\sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}},$$

$$\cos\alpha = \frac{X_2 - X_1}{\sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}}, \quad \cos\beta = +\sqrt{1 - \sin^2\beta} \quad \left(\beta < \frac{\pi}{2}\right),$$

$$\sin\gamma = -\sin\alpha\sin\beta + \cos\alpha\cos\beta, \quad \cos\gamma = -\sin\alpha\cos\beta - \cos\alpha\sin\beta,$$

$$X'_1 = X_1 + R_1\cos\gamma, \quad Y'_1 = Y_1 + R_1\sin\gamma, \quad X'_2 = X_2 + R_2\cos\gamma,$$

$$Y'_2 = Y_2 + R_2\sin\gamma.$$

In the case of an internal tangent (Fig. 5) the quantities needed for the calculation of X'_1, Y"1 and X'_2 and Y'_2 are:

$$\gamma_1 = \alpha - \beta + \frac{\pi}{2}, \quad \gamma_2 = \gamma_1 + \pi, \quad \sin\beta = \frac{R_1 + R_2}{\sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}};$$

$\sin\alpha$, $\cos\alpha$, and $\cos\beta$ are expressed as in the preceding case:

$$X_1' = X_1 + R_1 \cos \gamma_1,\ Y_1' = Y_1 + R_1 \sin \gamma_1,$$
$$X_2' = X_2 + R_1 \cos \gamma_2 = X_2 - R_2 \cos \gamma_1,$$
$$Y_2' = Y_2 + R_1 \sin \gamma_2 = Y_2 - R_2 \sin \gamma_1.$$

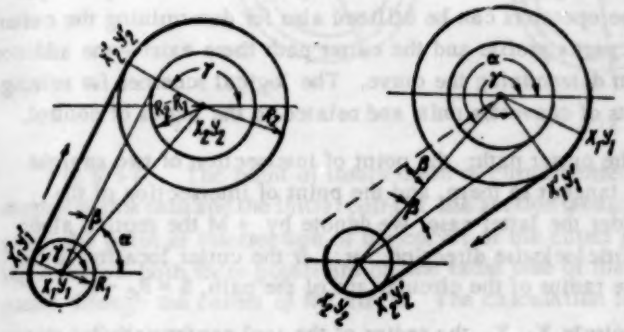


Fig. 4.

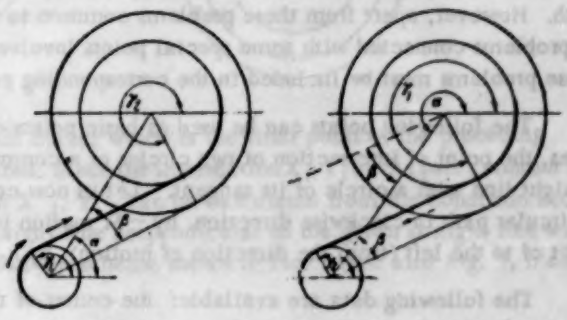


Fig. 5.

In calculating the external tangent and the motion along the contour in anticlockwise direction the fact is taken into account that $\gamma = [(\alpha - \beta) + \pi/2] + \pi$ and for the calculation of the internal tangent and motion along the contour in the anticlockwise direction $\gamma = [(\alpha + \beta) + \pi/2] + \pi$.

If the external tangent in the case of a clockwise motion is to be calculated:

$$\cos \beta \sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}$$
$$= \sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2 - (R_2 - R_1)^2} = U_1$$
$$\sin \gamma = \frac{-(Y_2 - Y_1)(R_2 - R_1) + (X_2 - X_1) U_1}{(Y_2 - Y_1)^2 + (X_2 - X_1)^2},$$
$$\cos \gamma = \frac{-(Y_2 - Y_1) U_1 - (X_2 - X_1)(R_2 - R_1)}{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}.$$

For the external tangent and during an anticlockwise motion the signs of sin γ and cos γ are changed.

For the calculation of the internal tangent with the motion in a clockwise direction:

$$\cos \beta \sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}$$
$$= \sqrt{(Y_2 - Y_1)^2 + (X_2 - X_1)^2 - (R_2 + R_1)^2} = U_2,$$
$$\sin \gamma_1 = \frac{-(X_2 - Y_1) U_2 + (Y_2 - Y_1)(R_2 + R_1)}{(Y_2 - Y_1)^2 + (X_2 - X_1)^2},$$
$$\cos \gamma_1 = \frac{-(Y_2 - Y_1) U_2 + (X_2 - X_1)(R_2 + R_1)}{(Y_2 - Y_1)^2 + (X_2 - X_1)^2}.$$

In the case of the internal tangent during anticlockwise motion the signs of sin γ and cos γ are reversed.

The described algorithm automatically discriminates between 96 different cases of contact between a straight line and two circles, taking into account the position of circles with respect to the origin of the coordinate system, and calculates the corresponding points of contact. The assumption that a point is a circle with radius R = 0 adds another 48 cases to the above number.

The conversion of a point into standard form means the determination of its Cartesian coordinates X and Y; the transformation of the circle is the finding of the coordinates of its center and the length of its radius. The conversion

algorithms of all methods of representing points and circles (see Table 3) are accumulated in converters of points and circles respectively.

3. Calculation of the Tool Center Curve

The problems involved in finding individual sections of the cutter center path curve are often the same which have been handled by converter units in converting command signals and in determining the elements of part profiles. By feeding into the APS the corresponding logics these operators can be utilized also for determining the cutter path. However, apart from these problems common to the part's profile and the cutter path there exist some additional problems connected with some special points involved in determining the curve. The logical schemes for solving these problems must be included in the corresponding points of converter units and related to the logics of control.

The following points can be used as basic points of the cutter path: the point of intersection of two straight lines, the point of intersection of two circles or a common tangent to them, and the point of intersection of the straight line with a circle or its tangent. Let us now consider the latter case. We denote by + M the motion along a circular path in clockwise direction, by − M motion in anticlockwise direction, by ± fr the cutter location to the right of to the left (from the direction of motion), by R_e the radius of the circular arc of the path, $S = R_e - R$.

The following data are available: the center of the circle X_3, Y_3, the radius of the tool center path R_e, sign M, sign fr, sign S, the direction of the straight section of the cutter's path, and the point of intersection of the straight line on the part's profile. Two cases are here possible.

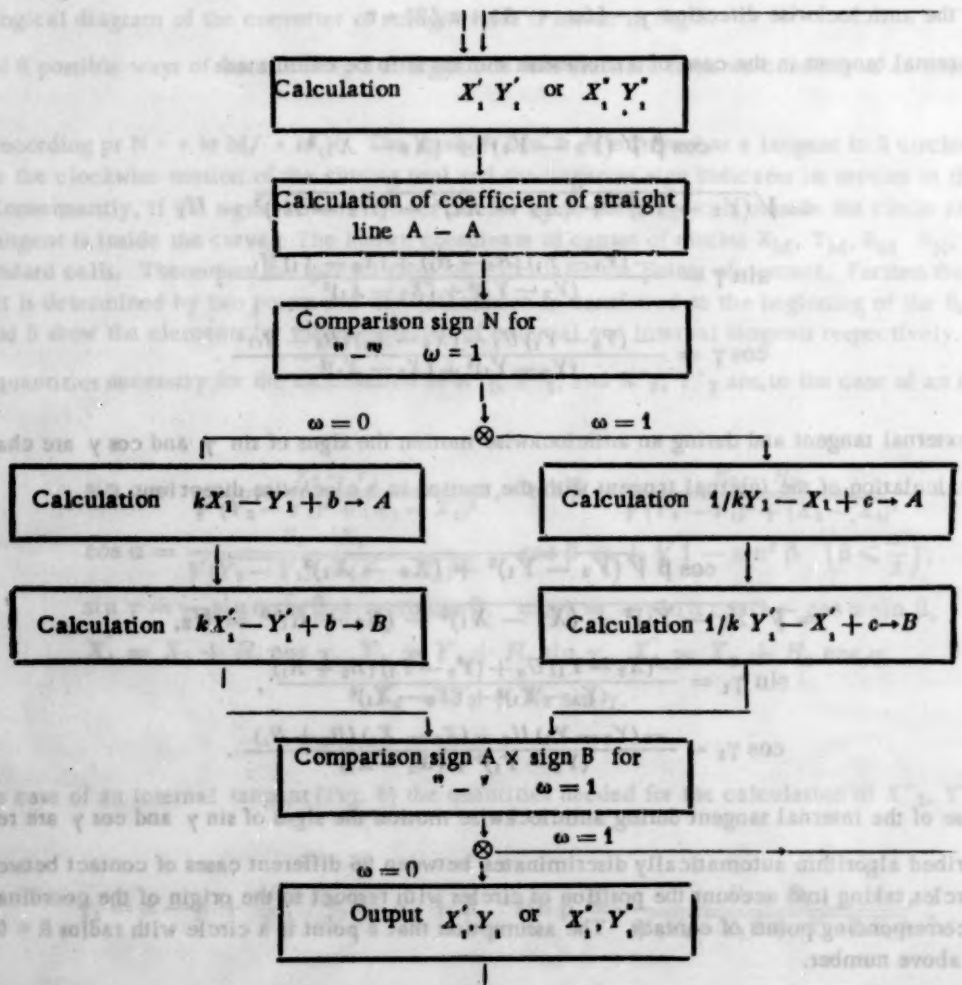

Calculation $X_1' Y_1'$ or $X_1'' Y_1''$
Calculation of coefficient of straight line A − A
Comparison sign N for ω = 1
ω = 0
ω = 1
Calculation $kX_1 - Y_1 + b \to A$
Calculation $1/kY_1 - X_1 + c \to A$
Calculation $kX_1' - Y_1' + b \to B$
Calculation $1/k\ Y_1' - X_1' + c \to B$
Comparison sign A × sign B for ω = 1
ω = 1
ω = 0
Output X_1', Y_1' or X_1'', Y_1''

Fig. 6.

Fig. 7

a) $S > 0$. The point of intersection of cutter center path curves which is the final point of the preceding section of this path and the initial point of the section is calculated. Since the intersection X_1, Y_1 on the part's contour is given, the point of intersection of the curves of the cutter path X'_1, Y'_1 can be calculated from the condition according to which both these points are on the same side of the straight line perpendicular to the given straight line which passes through the center of the circle. The calculation follows the scheme shown in Fig. 6 (see also Fig. 7, b and c, chain-dotted line).

b) $S < 0$. The distance l between the center of the circle and the straight line of the cutter center path is calculated first. Thereupon this distance is compared with the radius of the cutter center path arc.

If $l \leq R_c$ (Fig. 7, a and b, dotted line) then the point of intersection of both curves of the path is found as in the previous case.

If $l > R_c$ (Fig. 7, c, dotted line) then first the points K and L are found in which the straight line and the circle of the cutter's path contact the circle whose center is the point of intersection of the corresponding lines on the contour with the radius R_c. In this case the algorithm of intersection or of the touching of two circles is used.

4. Control Unit

The definition of the control unit was given in the first section. We described there also the function of two CU units during the operation of the APS – it is the decoding of the command signal and of the geometrical meaning of the signal. In this section we consider the functions of the CU in controlling the calculation process within the automatic programming system and characterize the principal moments in the operation of the logics of the APS.

For controlling the automatic programming system a certain amount of information must be contained in standard cells of the rapid-access (or constant) memory unit of the general-purpose computer; this information varies from cycle to cycle. By cycle we mean a sequence of logical and arithmetical operations which begins with the reading of command signals with the CP and ends with the reading of the next command signal of the CP. During each cycle in the operation of the APS (with the exception of the initial and final cycles) the standard memory cells contain and process the following information:

the command signal read at the beginning of the cycle;

the command signal read in the preceding cycle which started with "po", "ot", or "do";

the command signal read at the beginning of the cycle "p," corresponds to + 1, and at the beginning of the cycle "ot" and "lo" corresponds to – 1;

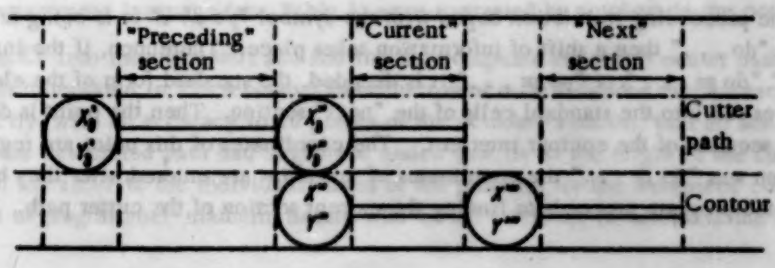


Fig. 8.

the command symbol read in the preceeding cycle;

the information on the position of the cutter center with respect to the contour of the part (to the right + 1, to the left − 1);

the information on the "current" section of part's profile (see Fig. 8);

the information on the "next" section of the part's profile.

the information on the "current" section of the cutting tool center path, the information on the "preceeding" section of the cutter path

the coordinates of the point of intersection of the "current" and "preceeding" sections of the profile X'' and Y'';

the coordinates of the point of intersection of the "current" and "next" sections of the profile X''' and Y''';

the coordinates of the point of intersection of the "preceeding" and "last-but-one" sections of the cutter path X'_e and Y'_e.

the coordinates of the point of intersection of the "preceeding" and "current" sections of the cutter path X''_e and Y''_e.

If the "preceeding" (or "current" or "next") section is a circular arc the information on the direction of motion on these sections is contained in the corresponding cells. Here + 1 denotes the rotation in clockwise direction and − 1 roatation in the anticlockwise direction.

If the "preceeding" or "current" section of the cutter path is a circular arc then the corresponding cells contain sign $(R'_1 - R_1)$ and sign $(R'_2 - R_2)$ respectively.

For sections in standard cells the following data are retained:

a) the symbol indicating the geometry of the section (circle or straight line);

b) data on the standard form of the section;

c) in the case of the circle the direction of rotation of the cutter on it; and

d) the symbol S_k with k = 1, 2 where S_1 = sign $(R_{e1} - R_1)$, S_2 = sign $(R_{e2} - R_2)$ which will be required for determining the path characteristics between the individual sections of the cutter path.

The position of the data in the memory unit of the computor is shown schematically in Fig. 8. In this illustration the data travel from right to left.

Let us now consider a cycle of operation, other than the initial or final cycle, where the type of cutting does not change (no transition from the external to internal milling, and vice versa, takes place). In this case all cells described above contain information. The cycle begins by reading the instruction from the CP tape. Let us assume that the read instruction begins with words "do . . ." or "po . . ."

If the instruction begins with the symbol "po . . ." and the preceeding instruction was " do kr . . ." or do pr . . ." then the read and the preceeding geometrical elements are compared. In the case of an agreement, the last instruction was an excessive information in CP and the APS proceeds to reading the next instruction. In the case of a disagreement the information is decoded and the standard form of the corresponding section calculated. The calculated figures are entered into the standard cells of the "next" section of the part's profile whereupon the APS turns to the reading of the next instruction. The procedure is the same if the preceeding command signal was "do tk . . ."

In the case when the preceeding instruction begun with the symbol "po . . ." or is being read at the given moment with the symbol "do . . ." then a shift of information takes place. Thereupon, if the information read was "do kr . . .", "po kr . . .", "do pr . . . ", or "po pr . . ." it is decoded, the standard form of the element determined and the obtained quantities fed into the standard cells of the "next" section. Then the point is determined in which the "next" and "current" sections of the contour intersect. The coordinates of this point are registered in cells X''', Y'''. If the read instruction was "do tk . . ." the coordinates of this point are entered, after they have been calculated into cells X''', Y'''. Now the program proceeds to finding the current section of the cutter path. The process depends

on whether an internal or an external surface is milled or whether the center of the cutter moves on the work's contour (the instruction is given at the beginning of CP). If there is no need to determine the cutter center path ("frO") then the data of the "current" section of the contour are fed directly into the cells of the "current" section of the curve of the cutter path. If this is a circle it is also necessary to remember + 1. If, however, the signal "fr" or "fr + " is in operation the cutter path must be calculated and the quantities obtained fed into the cells of the "current" section of the cutter path curve. In the case of a circle S_1 or S_2 is determined.

It is necessary to calculate the points of intersection of the preceeding and the current sections of the cutter path and to feed the corresponding coordinates into the cell X''_e, Y''_e. The cycle is completed by the calculation of differences and the typing of the necessary information.

After reading the signal "ot tk . . ." on the tape the geometrical meaning of the command signal is decoded and the coordinate of the point calculated and fed into cells X''', Y'''. Thereupon the new instruction is read from the CP tape. In the subsequent cycles, after shifting the information and the finding of new X''', Y''', X'' is transformed to X''_e, Y'' to Y''_e and the data of the "current" section of the contour are fed into the cells of the current section of the cutter path. Now the next instruction can be read from the tape.

For reading from the tape of the instruction "stop" shifts the information, transfers X'' to X''_e and Y'' to Y''_e, and also initiates the calculation of differences and typing and stops the computer.

The calculation of the cutter center path curve for a section of the part's profile is determined by the instruction which indicates the position of the tool center with respect to the part, i.e., cutter to the left, to the right, or on the profile; "fr ξ" (ξ = +, −, 0) (the radius of the cutter radius is introduced together with the numerical data). The instruction "fr ξ" -is operative during the entire program up to the next identical instruction. If the last instruction before the current "fr ξ" started with the symbol "po. . ." then the cutter path for the "next" section must be calculated according to the preceeding "fr ξ". For this reason the calculation of the cutter path according to the next "fr ξ" begins not before a double shift of information takes place when the standard cells of the "current" sections contains the data on the section punched on the CP tape following the signal "frξ". If the preceeding "fr ξ" signal started from the left,"do . . .",the standard cells of the "next" section contain (or receive during the next cycle not accompanied by the shift of information) the data determined by the instruction contained in the CP after "fr ξ". Consequently, the cutter path is calculated, in accordance with the new "fr ξ" instruction after the information has been shifted once.

5. Experimental Testing of APS

The basic part of the developed APS was programmed for a definite workpiece on a M-3 computer. The amount of information in the program was determined by the capacity of the rapid-access memory of the M-3 computer. The volume of the main parts of the program is shown in Table 4.

TABLE 4

Unit	No. of cells
Command of signal decoding unit and the decoding unit of the geometrical meaning of the signal	46
Control unit	214
Converter unit	510
Constants	41
Path curves	80

Since there were not special input devices for tansforming the language program into the binary code the letters and symbols of the programming language (see Table 1) were expressed by numbers in the octonary system.

After feeding the CP into the input unit and starting the computer the cutter center path curve for the part illustrated in Fig. 1 was calculated in 225 seconds. Examination of results showed that all sections of this curve had been calculated correctly (with an accuracy to 10 places of the octonary system), that no accumulation of errors had taken place, and that the calculated path had begun and ended exactly at the origin of the coordinate system. Table 5 shows the time taken to calculate the individual stages of the program for the workpiece concerned. The same problem was presented to programmer-mathematicians who took 7 - 8 hours to calculate the basic points of the cutter

TABLE 5

Program stage	Time in min.
Compiling the language program	20 - 25
Punching the language-program tape	10 - 12
Checking the tape	12 - 15
Feeding the designation group	8
Calculation on the M-3 computer	4
Total	54 - 64

path, using ordinary keyboard computers. The test shows that the use of the APS in calculating the basic points of the cutter path for a relatively simple work can speed up the calculation process 7 - 8 times. The gain greatly increases with more intricate parts.

Using more rapidly operated computers equipped with ferrite memory (e.g., "Minsk-1" and photo-electrical input units the calculation time can be reduced further to 1/60 which will reduce the total time required for programming by a further 20-25%.

SUMMARY

1. The automatic programming system eliminates the need to calculate the geometrical elements of the workpiece not given numerically and to set up the algorithm for their calculation; these functions of the program operator are transferred to a general-purpose computer.

2. The unknown geometrical elements of the contour can easily be determined by means of elements given numerically and by those determined earlier. The determined elements are recorded in terms of simple expressions of the programming language.

3. The preliminary calculation of the tool center path is obviated altogether; it is effected automatically by the general-purpose computer. The setting up of algorithms for calculating the cutter path curve is reduced to a simple description of the path by the controlling program written in the programming language.

4. Since the entire program prepared by a programmer is in the mnemonic language, the possibility of an error in this program is reduced to a minimum.

5. Since the programmer performs no calculations and the program is written in the programming language the time needed for its preparation is reduced to a fraction of the usual time.

All these considerations lead to two general conclusions.

Firstly, the automatic programming system makes possible a partial automation of the first stage of the programming process thus reducing the amount of mental work of the human operator.

Secondly, it eliminates the need to prepare accurate engineering drawings. These can be replaced by hand sketches, which considerably simplifies the work of the designer.

The author is greatly indebted to V. A. Kotel'nikov for his valuable suggestions.

LITERATURE CITED

1. Kitov A. I. and Krinitskii N. A., Electronic Digital Computers and Their Application. Fizmatgiz, 1959.
2. Chernyshev A. V. and Yakhin A. B., Automation of Machining on Metal Cutting Machine Tools by the Use of Program Control. Mashgiz, 1959.
3. Siegel A. Augomatic Programming of Numerically-Controlled Machine Tools. Control Engineering, No. 10, October, 1956.
4. Ross D. T. Automatic Programming. Aircraft Production, vol. 20, No. 5, May, 1958.

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

AUTOMATIC THICKNESS CONTROL OF COLD - ROLLED STRIP, BASED ON EDGE - THICKNESS MEASUREMENTS

Ch'ü Shou-tê (Peiping)

Translated from Avtomatika i Telemekhanika Vol. 22, No. 8,
pp. 1055-1060, August,1961
Original article submitted November 29, 1960

A new system of automatically controlling the thickness of cold-rolled steel strip, using as the basis the variation of the edge thickness, is described. The general design and the results of testing the system on an electronic model are given.

1. Problem Setup

The existing systems of automatic control fail to ensure an adequate uniformity of the gauge of the cold strip rolled on modern high-speed mills. Therefore a further improvement of the present gauge-control systems and the development of new types has become an urgent problem for the rolled metal industry.

The main cause of deviations of the thickness of cold-rolled strip from its nominal value is that the component parts of the mill are not perfectly rigid. Their elastic deformation varies, due to fluctuations of the forces acting on the rollers during the rolling process. The actual value of this deformation is commensurate with the thickness of the strip. This results in considerable variations in the thickness of the final product.

The forces acting on the rollers depend on a number of factors. The most important of these are the properties of the material being rolled and the rolling conditions. The greatest effect on the rolling accuracy is caused by the variation of the thickness of the edge of the original strip due to the "inherent" differences in the thickness of the hot-rolled strip, which is due to the eccentricity of the rollers and the variations of the forces in the mill on which it was produced. The main source of disturbances in automatic systems controlling the strip thickness during cold rolling are the variations of the edge thickness [1].

In the existing automatic thickness control units for cold-rolled strip the automatic control action is obtained by measuring the deviations of the thickness of the rolled strip which is measured by contact or contactless thickness guages situated after the stand. The timelag between the measurement and the reduction process caused by such arrangment of the thickness guage and the considerable inertia of the screwdown mechanism in such a system, makes it possible to obtain a high quality of regulation.

But such a system can be based on the measurement of the edge of the strip before the section in question enters the stand. The gain in time obtained by installing the edge thickness gauging unit some distance in front of the stand makes it possible to eliminate the unfavorable effect of the inertia of the screw-down gear, and this improves the quality of regulation. The designs of automatic control systems operating with automatic correction of the parameters of the compensating devices are considered in [1 - 3].

In the automatic thickness control system developed by the author at the Institute of Automation and Telemechanics at the Academy of Sciences, USSR, the controlling action on the screw-down gear is regulated by a thickness gauge arranged in front of the stand. The system is intended for controlling the thickness of the strip rolled on a reversing cold-rolling mill, or on the first stand of a continuous rolling mill in which the thickness is varied by operating the screw-down gear while the tension of the strip is held constand by means of a quick-acting automatic device. The deviation of the strip thickness from its nominal value is expressed by the equation

$$\Delta h = \frac{\partial h}{\partial H}\Delta H + \frac{\partial h}{\partial S}\Delta S, \qquad (1)$$

where ΔH is the deviation from its normal value of the thickness of the strip entering the stand and ΔS is the displacement of the upper roller from its normal position.

In order to make the deviation Δh of the strip thickness zero when the edge thickness $\Delta H \neq 0$, the upper roller must be moved through a distance ΔS. The value of the coefficient characterizing the ratio $\Delta S / \Delta H$ depends on a number of factors and varies continuously during the rolling process. In the automatic control system being considered a logical action relay system is provided to maintain this coefficient constant after comparing the deviations of the strip with the thickness of the edge in front of the roller.

2. The General Principle of the Gauge Control System

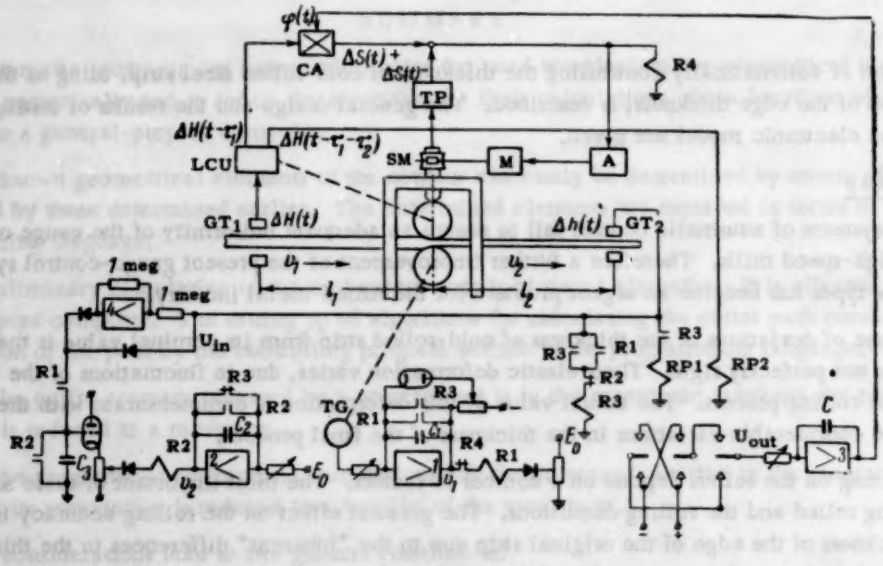


Fig. 1.

Figure 1 shows diagrammatically the automatic control system being described. The deviations ΔH of the edge thickness measured by the thickness gauge GT_1 are passed through the lag control unit LCU to the amplifier CA, whose amplification factor is variable. The drive of the LCU is synchronized with the main drive of the stand; the indications of the thickness gauge GT_1 are therefore fed into the amplifier CA after a periof of time τ_1, during which the strip travels from the point GT_1 where it was measured, to the rollers; $\tau_1 = l_1 / v_1$ there l_1 is the distance between the thickness gauge GT_1 and the axes of the rollers, and v_1 is the velocity of the strip entering the stand. Thus, the control action is passed on to the screw-down mechanism at the moment when the section of the strip with measured thickness enters the roller path. In fact, however, because of the acceleration of the screw-down mechanism, the time lag of the LCU is made slightly smaller than the time taken by the strip to move from GT_1 to the rollers.

The signal ΔS received by the servomechanism of the screw-down gear from the amplifier CA is compared with the signal $\Delta S'$ which shows the actual motion of the rollers as determined by the position transducer TP. The difference $\Delta S - \Delta S'$ controls the operation of the screw-down mechanism motors (through the amplifier A).

The amplifier CA provides the proportionality factor between the strip thickness ΔH, and the displacement of rollers ΔS needed to eliminate it. The amplification factor of the CA amplifier is proportional to the integral

$$\varphi(t) = k \int_0^t \operatorname{sgn} \Delta H (\xi - \tau_1 - \tau_2) \, \Delta h (\xi) \, d\xi + \varphi(0), \tag{2}$$

where τ_2 is the time taken by the strip to cover the distance between the rollers and the thickness gauge GT_2, installed after the stand which measures the thickness of the rolled strip, i.e., $\tau_2 = l_2 / v_2$ where l_2 is the distance between the roller axis and the thickness gauge GT_2, and v_2 is the velocity of the strip leaving the pass.

If the deviation of the thickness of the rolled strips has the same sign as the deviation of the edge thickness, the amplification factor of the CA amplifier is too small, i.e.,the compensation of the system is inadequate. In this case the amplification factor should be increased, with the result that the integral in (2) increases. If the sign of the deviation of the rolled strip from the required value is different from that of the strip entering the rollers, then the amplification factor of CA exceeds the requied value, i.e., the system is overcompensated. In this case the integral should be reduced. The amplification factor of the CA amplifier is regulated by means of an impulse unit. This can be effected by means of a logical-action relay circuit and integrator, on the basis of the comparison of indications of both thickness gauges, with the time lag taken into account.

The impulse integrator unit consists of two electronic integrators, a relay unit, a rectifier, and a capacitor; it determines the duration of the regulating action on the amplifier CA in accordance with the strip thickness measured by the GT_2 gauge and also the length of the intervals between these actions on the basis the voltage produced by the tachometric generator TG, which is positively connected with the main drive.

During the pause the relay R3 is deenergized. The input of integrator 1 receives voltage proportional to the rolling speed, while the output voltage monotonically increases until its value exceeds the standard voltage of the source E_0. The input resistance of integrator 1 is set so that after an interval equal to the time lag of the thickness measurement of the rolled strip τ_2, the output voltage of the integrator becomes equal to the standard voltage and operates relay R1. The contacts of relay R1 bypass the capacitor C_1 for a short period of time, with the result that the output voltage of integrator 1 decreases to zero. At the same time the winding of relay R3 becomes connected to the thickness gauge GT_2 by the other pair of contacts of relay R1. If the thickness of the rolled strip is the same as the required value there will be no voltage at the thickness gauge output: consequently, the amplification factor of the CA amplifier has been selected correctly and requires no adjustment. In this case relay R3 will not operate. After a brief bypassing of the capacitor C_1 by the contacts of relay R_1 the cycle of integration by the integrator of voltage generated by the tachometric generator TG is repeated. If the amplification factor of amplifier CA has been selected incorrectly the thickness of the rolled strip differs from its normal value and produces a voltage at the output of the thickness gauge TG_2. The closing of the contacts of relay R1 operates relay R3 and blocks itself. Then the winding of the polarized relay RP2 becomes connected to the output of the thickness gauge GT_2, and the logical action relay circuit is activated either increasing or reducing the amplification factor of amplifier CA by means of integrator 3. At the same time the feedback link of integrator 1 is bypassed by the normally open contacts of relay R3, and the integrator is disconnected.

During the operation of relay R1 the absolute value of the output voltage of the thickness gauge GT_2 passed through the third pair of contacts of relay R1 is remembered by capacitor C_3. During the pause capacitor C_2 of the feedback link of integrator 2 is bypassed by the normally closed contacts of relayR3, and the output voltage of the integrator is zero. During the operation of relay R3, integrator 2 begins integration and its output voltage U_2 shows a linear increase. After a period of time, which can be adjusted by means of the input resistance of integrator 2, the output voltage reaches a value taken from capacitor C_3 by means of a cathode repeater. This operates relay R2, which disconnects the winding of relay R3 and this, in turn, releases the polarized relay RP2. Thus, the time of operation of the logical-action relay circuit is determined by the operation time of integrator 2, which is proportional to the deviation of the strip thickness measured after leaving the stand at selected moments of time. After the relay R3 has been released the feedback link of integrator 2 is again bypassed, and its output voltage rapidly decreases to zero. At this time integrator 1 begins the next integration cycle, i.e., a new pause. Consequently, electronic integrators 1 and 2 operate in turn and determine the times of connecting and disconnecting and adjustment of amplifier CA.

3. Some Results of Experimental Testing of the System on an Electronic Model

The testing of the system described above on an electronic model showed that the uniformity of thickness of the rolled strip which can be obtained with the system described depends mainly on the speed of response of the screw-down mechanism of the stand, on the rate with which the parameters of the stand change,and on the speed with which the amplification factor of the CA amplifier can be changed.

Figure 2 shows an oscillogram which illustrates the variation of the input voltage of the unit U_{in}; this voltage is proportional to the deviation Δh of the thickness of the rolled strip - the latter was given as a sinusoidal curve. When the system is in operation the output of the impulse unit shows a sequence of square pulses U_{out}, which are passed on to integrator 3 whose task it is to adjust the amplification factor of amplifier CA. The oscillogram also shows the sawtooth curves representing the operation of integrator 1 (U_1) and 2 (U_2).

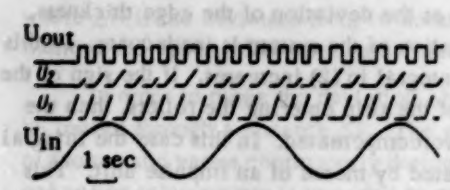


Fig. 2.

Owing to the inertia of the screw-down mechanism the system cannot alter sudden variations of strip thickness. This results in a deviation of the thickness of the rolled strip during the transient period of the screw-down gear. Such deviations are independent of the amplification factor of amplifier CA and are observed even when the amplification factor of CA is set correctly. After the end of the transient period all low-frequency variations of edge thickness are compensated. The device which adjusts the amplification of amplifier CA must not react to these variations of thickness of the rolled strip. To achieve this, relay R4 is placed at the servomechanism input whose normally closed contacts are connected to the output of integrator 1 (Fig. 1). The dead zone of relay R4 is made larger than the corresponding zone of the system controlling the lifting and lowering of the rollers. During the transient process the unbalance voltage of the servomechnism exceeds the dead zone of the relay R4, and its normally closed contacts open. This does not operate relays R1 and R3 while the pulse unit remains at rest during the pause. Therefore, the amplification factor of CA as set before the transient process of the servomechanism remains unchanged. In order to ensure that, during a prolonged pause, the output voltage of integrator 1 does not exceed the permissible value, a diode limiter is provided in the feedback circuit of the integrator.

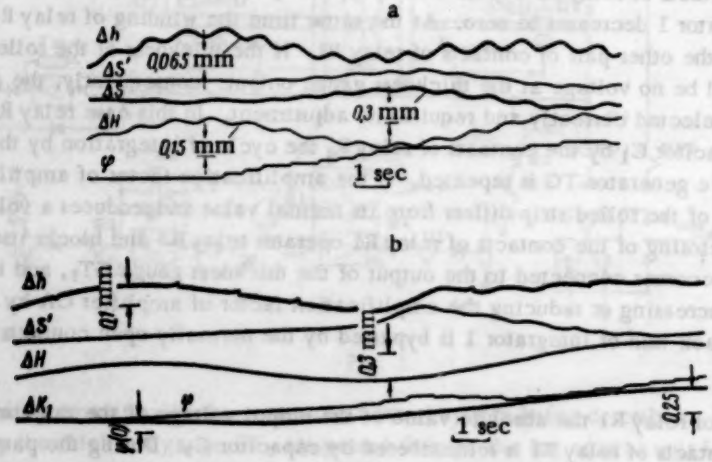


Fig. 3.

The results of this experimental investigation of the thickness control system on an electronic model representing a reversing cold-rolling mill for non-ferrous metals showed that this system should improve the uniformity of strip thickness more than the existing systems. It eliminates all slowly changing variations of thickness but leaves unaffected high-frequency changes (with a frequency above 0.2 cps), as shown on the oscillogram (Fig. 3a). The system is stable in operation despite the disturbances, and ensures a high quality of regulation when the amplification factor of the mill stand varies in accordance with a sinusoidal law, with an amplitude corresponding to 20% of the average value and with a frequency of 0.05 cps (Fig. 3b). The positive results obtained in these tests support the satisfactory operation of this control system.

LITERATURE CITED

1. Chelyustikin, A. B. "Automatic Control of strip profile on continuous cold-rolling mills". A paper presented to the All-Union Conference on the Automation of Production Processes. Scientific-Technical Society of Ferrous Metallurgy, 1956.
2. Chelyustkin, A. B. and Ivanov, V. A. "A self-adjusting automatic control system for electrical tube-welding machines", Elektrichestvo, No. 2, 1960.
3. Doganovskii, S. A. and Fel'dbaum, A. A. "An investigation on an electronic model of the degree to which variations of edge thickness during rolling of strip have been compensated". Avtomatika i telemekhanika, 20, No. 2, 1959.

SOME APPLICATIONS OF CONTACT GRIDS*

Antonin Svoboda (Prague)

Director of the Institute of Mathematical Machines, Academy of Sciences,
Prague, Czechoslovakia

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 1061-1070, August, 1961
Original article submitted February 2, 1961

At the present time there is a tendency to solve all problems of switching circuits by using electronic computers. Where such a facility is not available, or in the case of an easier problem, graphs, tables, and maps are used. We try to prove here that even a very primitive mechanization of computing in Boolean algebra is desirable. As an example, we show how to find a minimal form of a given Boolean function with the aid of contact grids, directional grids, and special symbols for integers. The theoretical background of this problem has been well treated in the literature. In our discussion we shall use the language used in the paper of Urbano and Mueller [1]. (This paper also contains a list of other theoretical papers.)

Contact Grids

Contact grids were described by A. Svoboda [2, 3] as graphical-mechanical aids for computing in Boolean algebra. For a detailed description of contact grids we refer the reader to the original paper. Here we shall simplify our task by fixing our attention on contact grids handling functions of six independent variables,

$$x_0 = x,\ x_1 = y,\ x_2 = z,\ x_3 = X,\ x_4 = Y,\ x_5 = Z.$$

The 6-cube is represented in the form of a map as shown in Fig. 1. A given Boolean function f(ZYXzyx) is projected on the map by marking with a dot the states (vertices) in which $f_i = 1$. The function mapped in Fig. 1 has

$$f(110101) = f(G5) = 0,\ f(001010) = f(B2) = 1.$$

An assortment of contact grids for six variables Z, Y, X, z, y, x is drawn in Fig. 2. The grids are made from paper, cardboard, or plastic material. By rotating them about the axes marked in the figure we obtain grids representing $\overline{Z} = x_5$, $\overline{Y} = x_4, \ldots, \overline{x} = x_0$.

If all 6 (= n) grids are superposed, as sketched in Fig. 3, only one little square is transparent. The location of this window depends on the selection of the grids. As shown in the figure the window corresponds to the block B2 of the map, and the grids selected are $\overline{Z}$, $\overline{Y}$, X, $\overline{z}$, y, $\overline{x}$. The set of grids can be taken as a model of the Boolean term $\overline{Z} \cdot \overline{Y} \cdot X \cdot \overline{z} \cdot y \cdot \overline{x}$, the window indicating the position on the map of the corresponding point (vertex) B2 of the 6-cube.

In the language used by Urbano and Mueller we regard the unit n-cube as being made up of cells:

0-cell or vertex—a point	k = 0
1-cell—line segment	k = 1
2-cell—quadrilateral	k = 2
3-cell—hexahedron	k = 3
k-cell	k

* Copyright, 1959, Harvard University Press. This article originally appeared in PROCEEDINGS OF AN INTERNATIONAL SYMPOSIUM ON THE THEORY OF SWITCHING, VOL. I, and has been reproduced here by permission of Harvard University Press.

The vertices of the 6-cube correspond to the terms of the canonical form of the given Boolean function: $+\dot{Z}\cdot\dot{Y}\cdot\dot{X}\cdot\dot{z}\cdot\dot{y}\cdot\dot{x}$, where the dot above a symbol means the choice between a variable (say y) or its complement (say $\bar{y}$). Any selection of 6 contact grids represents a model of a vertex.

A k-cell corresponds to a term with n - k variables. Any selection of n - k contact grids is a model of a k-cell. There are 2^k windows in the model corresponding to the 2^k vertices, the 0-faces, of the k-cell.

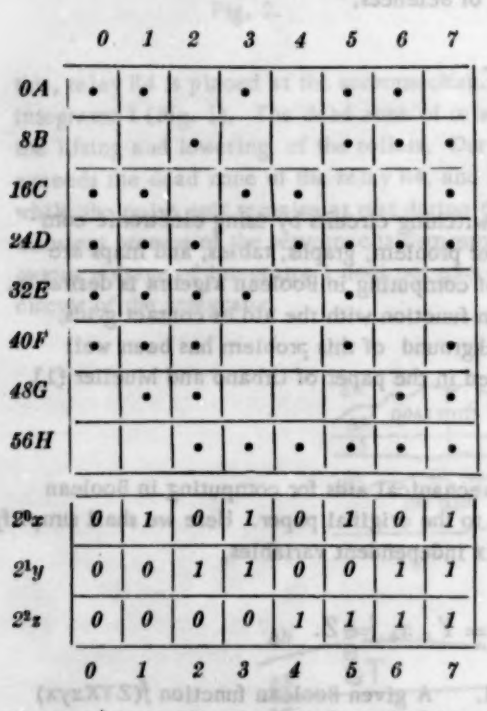


Fig. 1. A map of a given Boolean function $f(Z,Y,X,z,y,x)$.

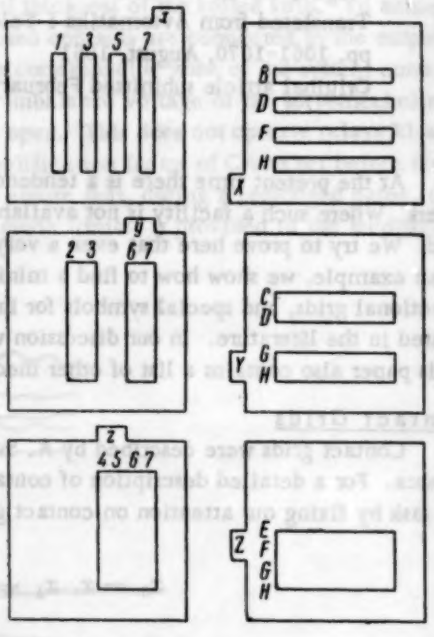


Fig. 2. Contact grids for six independent variables Z, Y, X, z, y, x.

Example 1. The function mapped in Fig. 1 is a 6-dimensional body immersed in a unit 6-cube. The vertices of the body are represented by dots. A dot being marked in the block B2, the value of the function $f(Z, Y, X, z, y, x) = f(001\ 010) = 1$. The canonical form therefore includes the term $\bar{Z}\bar{Y}X\bar{z}y\bar{x}$.

Example 2. Find all cells incident with the vertex B2 in the body defined in Example 1. First we make a model of the vertex B2 as shown in Fig. 3. By placing it on the map of the given function (Fig. 1) we check the model: a dot must be seen through the window. It is clear that if we take away from our model any number of grids the window B2 will remain transparent. By taking away k grids we obtain a model of a k-cell incident with B2. Such a model will have 2^k transparent windows, corresponding to the positions of the vertices of the corresponding cell. Looking at the map of the function through the model we see at once whether the cell belongs to the body of the function or not. If so, we see dots (symbols for the value 1) in all windows of the model. The search can be done systematically starting with all 6 (=n) 1-cells represented by sets of 5(=n−1) grids. The favorable cases are:

The original combination:	$\bar{Z}\bar{Y}X\bar{z}y\bar{x}$	0-cell	B2
The 5-grid combinations:	$\bar{Z}\bar{Y}Xy\bar{x}$	1-cell	B26
	$\bar{Z}X\bar{z}y\bar{x}$	1-cell	BD2
	$\bar{Y}X\bar{z}y\bar{x}$	1-cell	BF2
The 4-grid combinations:	$\bar{Z}Xy\bar{x}$	2-cell	BD26
	$X\bar{z}y\bar{x}$	2-cell	BDFH2

The favorable 5-grid combinations were found from a favorable 6-grid combination. In the same way the favorable 4-grid combinations are found from favorable 5-grid combinations. The 2-cell BD26 is obtained from the 1-cell B26 by taking away the grid Y (Fig. 4).

Remark. The reader will note that the symbols employed here, for example, B2, B26, BDFH2, represent Boolean forms, for example, , the symbol for intersection falling always between the letters and the numbers.

Basic Cells and Stars

The symbolism introduced in the last example makes it quite easy to enumerate all basic cells incident with a given vertex. The cell B26 is not basic as it is included in the cell BD26 (B is in BD, 26 is in 26). In the same way, we eliminate B2, BD2, BF2 in the last example. The cells BD26, BDFH2 which remain are basic. They form

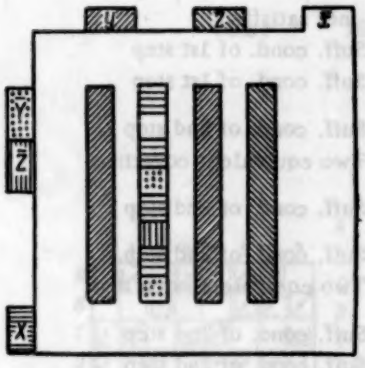


Fig. 3. Model of the Boolean term $\overline{Z} \cdot \overline{Y} \cdot X \cdot \overline{z} \cdot y \cdot \overline{x}$ (the block B2 on the map).

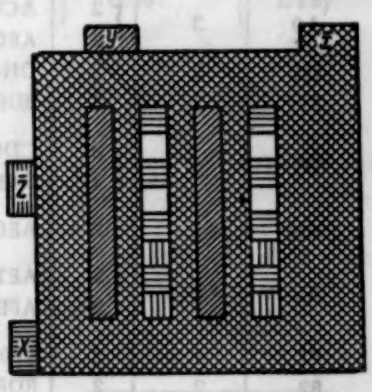


Fig. 4. Model of the 2-cell BD26 (a 4-grid combination $\overline{Z} \cdot X \cdot y \cdot \overline{x}$).

the basic star of the vertex B2 in the sense of Urbano and Mueller. We could write down a table of basic stars, find the essential stars, enumerate the irredundant coverings, and conclude with the selection of minimal coverings. We shall show a method for solving the same problem using other graphical and mechanical aids.

Directional Grids

Starting from a given vertex of the unit n-cube there are n 1-cells (line segments) emanating in n different directions. We shall describe a simple aid, in the form of grids which indicate on the map all possible 1-cells incident with a given vertex.

Directional grids are models of 2^n bodies described as follows (Fig. 5): The directional grid 0 is the model of a body possessing as vertices all points 0 (that is, A0, B0, . . . , H0) and all end-points of 1-cells (line segments) emanating from points 0 in all directions.

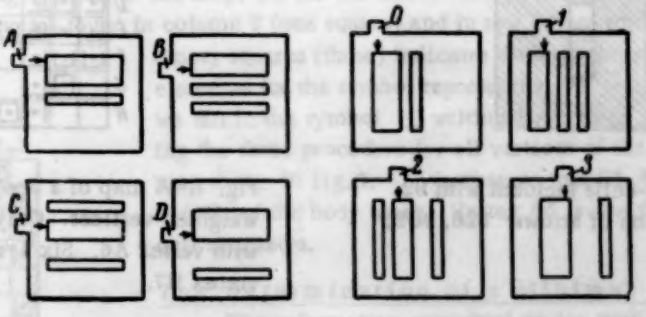


Fig. 5. Directional grids for six variables. By rotating the grid 0 about the vertical axis we get the grid 7 (and so on).

In the same way we define the directional grids 1, 2, 3, . . . , 7, A, B, . . . , H. Thus the directional grid B is a model of the body possessing as vertices all points B (that is, B0, B1, B2, . . . , B7) and all end-points of 1-cells (line segments) emanating from points B in all directions. Only half of the grids need be made: grids 4, 5, 6, 7 are obtained by turning over grids 3, 2, 1, 0, and grids E, F, G, H are obtained by turning over grids D, C, B, A.

TABLE 1

Vertex	Weight	k	Cell	Term	Remarks
A6	1	1	AB6	$\bar{Z}\bar{Y}zy\bar{x}$	Suff. cond. of 1st step
C4	2	2	CD04	$\bar{Z}\bar{Y}y\bar{x}$	Suff. cond. of 1st step
G1	2	2	ACEG1	$\bar{x}\bar{z}yx$	Suff. cond. of 1st step
H3	2	2	H2367	ZYXy	Suff. cond. of 1st step
A0	3	2	AC01		Suff. cond. of 1st step
		2	AE01		not satisfied
H4	3	3	DH4567	YXz	Suff. cond. of 1st step
F7	3	3	BDFH57	Xzx	Suff. cond. of 1st step
D1	0	2	CD01	$\bar{Z}Yz\bar{y}$	Suff. cond. of 2nd step
		2	D0145	or $\bar{Z}YX\bar{y}$	Two equivalent coverings
A0	1	2	AE01	$\bar{Y}\bar{X}z\bar{y}$	Suff. cond. of 2nd step
A5	1	2	AE15	$\bar{Y}\bar{X}yz$	Suff. cond. of 2nd step.
		2	ABEF5	or $\bar{Y}z\bar{y}x$	Two equivalent coverings
F1	1	2	EF01	$Z\bar{Y}z\bar{y}$	Suff. cond. of 2nd step
B2	2	2	BDFH2	$Z\bar{z}y\bar{x}$	Suff. cond. of 2nd step

To find all 1-cells emanating from a given vertex of a body corresponding to a Boolean function described by our map, we place suitable directional grids on the map and read the cells.

For instance, using grids B2, we obtain the configuration shown in Fig. 6. Looking along column 2 and row B, we read 1-cells: BD2, BF2, B26. Certain remarks concerning cells of higher order we leave to the reader. We point out that only three 1-cells (segments) connect vertex B2 with its neighbors: B6, D2, F2.

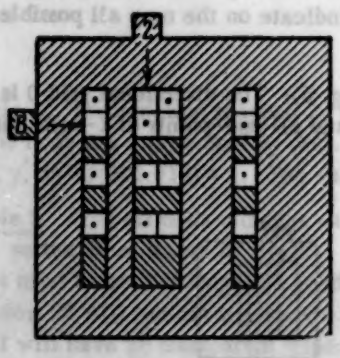


Fig. 6. Reading all 1-cells incident with B2. Reading in the direction of arrows: B26, BD2, BF2.

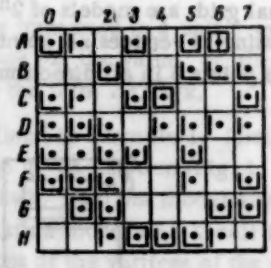


Fig. 7. A map of a Boolean function showing weighted vertices. Only one 1-cell is incident with vertex A6. Six 1-cells are incident with vertex H7.

Weighting Vertices

We take the number of 1-cells incident with a vertex as its weight W. On the map of the function we mark these weights by special symbols as indicated in the following table:

Weight W = 6	5	4	3	2	1	0
Symbols	.					

TABLE 2. First Trial Solution for the Function of the Example

K	Cell	Term	Number of operations
2	E0123	$\bar{Z}\bar{Y}\bar{X}\bar{z}$	4
2	AC13	$\bar{Z}\bar{X}\bar{z}x$	4
2	CDGH7	$Yzyx$	4
2	GH26	$ZY\bar{y}\bar{x}$	4
		Total	16

TABLE 3. Second Trial Solution for the Function of the Example

K	Cell	Term	Number of operations
2	EFGH2	$Z\bar{z}\bar{y}\bar{x}$	4
2	AE13	$\bar{Y}X\bar{z}x$	4
1	C37	$\bar{Z}\bar{Y}\bar{X}yx$	5
2	GH67	$ZYzy$	4
		Total	17

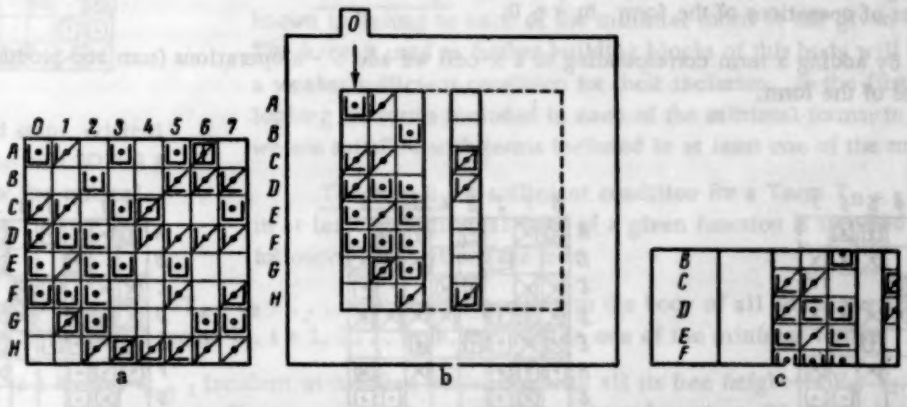


Fig. 8. a) Marking of the cells satisfying the sufficient condition for inclusion (Theorem 1); b) the application of the directional grid 0 to reduce free weights of vertices. We work in the rows C, D where a new bound vertex appears. Only weights of free vertices are reduced. Therefore no change occurs in the row C; c) the situation after the reduction of weights in rows C, D as the result of the application of the directional grid 0. (Added graphical elements are indicated by heavy lines.)

The symbols are composed so as to make graphical subtraction possible. To subtract m units from a weight we add m graphical elements to its symbol (the symbol for a weight W is composed of n−W line segments).

To plot the weighting symbols in the map, we use the directional grids as shown in Fig. 6. It is only necessary to count the empty squares found in column 2 (one square) and in row B (two squares). The total number of empty squares (three) indicates directly the number n−W of graphical elements for the symbol representing the weight of the vertex. In our case we fill in the symbol [·] writing down three graphical elements. By repeating the same procedure for all vertices of the given function we obtain the map shown in Fig. 7. We observe that E1, H7 are vertices at which six 1-cells of the body meet. Vertex A6 is one from which only a single 1-cell emanates.

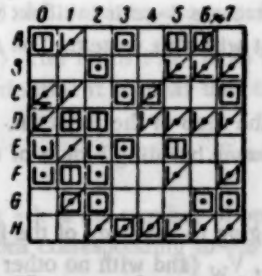


Fig. 9. The map after completion of the first methodic step. Free weights of free vertices have been reduced.

The Determination of a Minimal Form

The information contained on our type of map of a Boolean function can be used to find a minimal form of the function. We have two models of the function at our disposal: the abstract model represented by the body embedded in the unit n-cube, and the concrete model represented by the map. On the map we have additional information about the weights of vertices of the body. It is possible to visualize certain important features of the body by observing the patterns formed by vertices with regard to their weights.

A minimal form will be compiled from terms corresponding to cells used as building blocks for the body of the function. The construction of the body and the composition of a minimal form will proceed in parallel.

First step. We start by processing the information contained in the original map of the function, looking for terms which must be included in each of the possible minimal forms of the given function.

Theorem 1. A sufficient condition for inclusion of a term T in (any of) the minimal forms of a given function is the incidence of the corresponding k-cell t with a vertex V of weight k.

Proof. A k-cell has k 1-cells incident with any of its own vertices. All k 1-cells emanating from the vertex V therefore belong to the k-cell t incident with this vertex. For this reason all (k - 1)-cells, (k - 2)-cells, . . . , 0-cells incident with V belong to the k-cell t (being its faces). We have to use one of these cells as a building block for the body of the function. To get a minimal form only the k-cell t can be used. If we were to use a (k - s)-cell, s = 1, 2, . . . , k, the corresponding term would add n - (k - s) operations to the number of operations of the minimal form. Such a form would not be minimal, because by substitution of the k-cell for the (k-s)-cell we could reduce the total number of operations of the form by s > 0.

Remark. By adding a term corresponding to a k-cell we add n - k operations (sum and product) to the number of operations of the form.

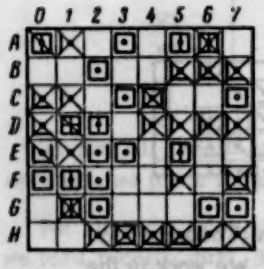


Fig. 10. Second step; critical vertices: D1, A0.

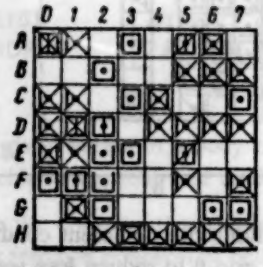


Fig. 11. Second step; critical vertex: A5.

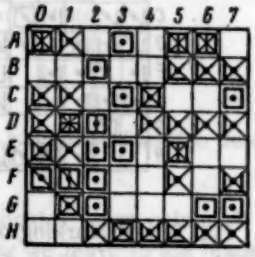


Fig. 12. Second step; critical vertex: F1.

Example 3. A Boolean function is given on the map in Fig. 7. Find the terms of the minimal form fulfilling the sufficient condition presented in Theorem 1. We always start with vertices of smallest weight; for vertices of weight 0 or 1 there is always a cell fulfilling the sufficient condition. During the following discussion we refer to Table 1.

The vertex A6 has weight 1. Therefore there is only one vertex with which it is connected through a 1-cell. Using contact grids or directional grids we find the 1-cell to be AB6 = $\bar{Z}\bar{Y}zy\bar{x}$.

The vertex C4 has weight 2. Using grids we find the 2-cell CD04 incident with C4. All minimal forms must include the corresponding term $\bar{Z}Y\bar{y}\bar{x}$.

Similar arguments apply to vertices G1, H3, H4, F7. In the last two cases there are 3-cells incident with vertices of weight 3. The vertex A0 is also of weight 3, but there is no 3-cell incident with this vertex. The sufficient condition is not fulfilled for any of the 2-cells incident with this vertex.

The elements of the minimal form compiled in the first step are plotted on the map of the function. Henceforth the vertices of the body of the function fall into two classes: bound (to the chosen building blocks of the body of the function) and free.

Definition. The free weight W of a vertex V_0 is the number of 1-cells belonging to the body of the Boolean function incident with the vertex V_0 and incident with W free vertices $V_1, V_2, \ldots, V_W$ (and with no other vertex). The vertices $V_1, V_2, \ldots, V_W$ are called free neighbors of the vertex V_0.

The first step includes the determination of free weights for all free vertices. When a building block is added in constructing the body of the function the free weights of some vertices are reduced. This reduction is marked down on the map quite easily, by virtue of the property of the symbols used. To reduce a weight by m units we

add m graphical elements to the corresponding symbol for the weight. The adjustment of the weights is mechanized by the use of directional grids.

Example 4. Complete the first methodic step using the results obtained in Example 3. First we mark down the cells satisfying the sufficient condition of inclusion (Fig. 8a). We then reduce the free weights of free vertices. Starting with the directional grid 0, we place it properly on the map (Fig. 8b) and in the column 0 look for vertices which have been bound. They are two: C0, D0. In the rows of these vertices we reduce the free weight of each free vertex by one by adding a graphical element to its symbol. The detail of Fig. 8c shows the situation immediately after the reduction of free weights by means of the grid 0.

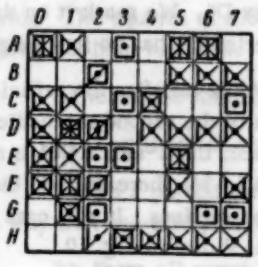


Fig. 13. Second step; critical vertex: B2. The cell BDFH2 is also a K-cell for the critical vertex D2.

In the same way we apply grids 1, 2, . . . , 7, A, . . . , H. The final result of the reduction is shown in Fig. 9.

Second step. The body of bound vertices built during the first step is known to belong to each of the minimal forms of the given Boolean function. The k-cells used as further building blocks of this body will be selected using a weaker sufficient condition for their inclusion. In the first step we were looking for terms included in each of the minimal forms; in the second step we are satisfied with terms included in at least one of the minimal forms.

Theorem 2. A sufficient condition for a Term T_{n+1} to be included in at least one minimal form of a given function is satisfied when all the following propositions are true:

1. The partial form $F_n = T_1 + T_2 + \ldots + T_n$ (corresponding to the body of all bound vertices built from cells $t_1, t_2, \ldots, t_n$) contains terms T_i, $i = 1, 2, \ldots, n$, included in one of the minimal forms;

2. There is a K-cell t_{n+1} incident with a free vertex V_0, with all its free neighbors $V_1, V_2, \ldots, V_W$, with free vertices $V_{W+1}, V_{W+2}, \ldots, V_{W+p}$, and with any number of bound vertices ($W \leq K$ being the free weight of the vertex V_0);

3. There is no k-cell incident with V_0 and with another free vertex not belonging to the set $V_0, V_1, \ldots, V_{W+p}$;

4. All k-cells incident with V_0 have $k \leq K$;

5. The term T_{n+1} corresponds to the K-cell t_{n+1}.

Proof. The body F_n of bound vertices does not include the free vertex V_0. There must therefore be at least one k-cell t'_{n+1} incident with V_0 which will produce the term T'_{n+1} of a specific minimal form F' (that is, one of the minimal forms). The form F' can be transformed into another form F by deleting the term T'_{n+1} (deleting the k-cell t'_{n+1}) and adding the term T_{n+1} (introducing the K-cell t_{n+1}). Then the following statements are true:

a) The vertices uncovered by deleting the k-cell t'_{n+1} include one or more of the vertices $V_0, V_1, \ldots, V_{W+p}$ and no others [cf. statement 3];

b) All vertices uncovered by deleting t'_{n+1} are covered by the K-cell [cf. statement 2];

c) The forms F, F' have the same combined sums S, S' of the number of + and · operations. The form F' being minimal it is clear that $S \geq S'$. On the other hand we can write

$$S = S' - (n - K) + (n - K) = S' + k - K \leqslant S',$$

where n - k is the number of operations corresponding to the deleted term T'_{n+1}, and n - K is the number of operations corresponding to the added term T_{n+1} [cf. statement 4]. We conclude that S = S'.

The form F being minimal, we see that at least one of the following two statements is true:

I) At least one minimal form F' exists which does not include the term T_{n+1} (corresponding to K-cell t_{n+1}); then the term T_{n+1} belongs to a minimal form F which may be constructed from F';

II) A minimal form F' which does not include T_{n+1} does not exist.

It has thus been proved that a K-cell satisfying the sufficient condition defined in Theorem 2 produces a term T_{n+1} belonging to at least one minimal form.

Example 5. Perform the second methodic step based on the results obtained in Example 4.

The free vertex D1 has free weight W = 0. There are no free neighbors. There are two K-cells: CD01, D0145 fulfilling the sufficient condition of Theorem 2. Both cells have K = 2, and each will produce a different minimal form (see Table 1). By including either of the terms ZYzy, ZYXy we bind the free vertex D1. We mark it on the map by crossing it out. This will not change the free weights on the map because the vertex D1 has no free neighbors.

The vertex A0 has free weight 1 and therefore one free neighbor E0. The 2-cell AE01 satisfies the sufficient condition of Theorem 2. Statement 1 of Theorem 2 remains valid after each repetition of the second methodic step. Statement 2 is true since W = 1, K = 2. Statements 3 and 4 check. For instance, the 2-cell AC01 is not a K-cell. It is incident with A0 and has k = 2 but it does not cover E0, the neighbor of A0. We therefore add the term $\overline{YX}\overline{zy}$, cross out the vertices A0, E0, and reduce the weights of the neighbors of these vertices. In this case the free weights of vertices F0, E2 are reduced by one. The appearance of the map is then as shown in Fig. 10.

Vertex F1 has weight 1. Vertex F0 is its only free neighbor. The 2-cell EF01 is not a K-cell. It is incident with F1 and with all its neighbors, but there is a cell EF15 incident with F1 and with free vertex E5 which is not covered by EF01 [cf. statement 3 of Theorem 2]. As it is impossible by the means at our disposal to make a decision for F1, we try another vertex.

Vertex A5 has free weight 1. There are two equivalent 2-cells AE15, ABEF5 satisfying the last sufficient condition. Each covers the free neighbor E5. There is no change of free weights on the map since vertex E5 has also the free weight 1, and therefore there cannot be any other free neighbors of either A5 or E5. The situation is shown in Fig. 11.

Now vertex F1 can be covered by a K-cell EF01 because vertex E5 is no longer free. We come to the situation in Fig. 12.

Vertex B2, having free weight 2, has free neighbors D2, F2. The 2-cell BDFH2 covers both neighbors and satisfies the last sufficient condition. The same is not true for the 2-cell BD26. After this step the situation on the map is as shown in Fig. 13.

Third step. Using the sufficient condition of Theorem 2 we come to the situation (illustrated in Fig. 13) where all of the criteria used previously fail. Then we start with some vertex of smallest possible free weight and try all possible coverings of its neighbors. For each covering we finish the compilation of the minimal form by repeating the second methodic step, count the number of operations introduced in each of the resulting forms, and decide which trial has led to a minimal form.

In case the second step, made after a trial, fails to cover all remaining free vertices, we have to make a second trial (repeat a third step) followed by another repetition of the second step. Generally each failure leads to a repetition of the third methodic step; this is connected with the necessity of investigating possibilities corresponding to all ramifications considered at all trials.

Example 6. Perform the third step using the results obtained in the preceding examples. The sufficient condition of Theorem 2 does not apply to any of the free vertices in Fig. 13. As all vertices possess the same weight, we start with an arbitrary one, say E2. Using 2-cell E0123 as a first trial covering for vertex E2 leads to the solution indicated in Table 2. Using 2-cell EFGH2 as a second trial covering leads to the solution indicated in Table 3. Both solutions are irredundant, and the solution in Fig. 13 is minimal (it has the smallest possible number of operations). We have derived one of the minimal forms of the Boolean function given on the map:

$$f = \overset{1}{\overline{Z}}\overline{Yzyx} + \overset{2}{\overline{Z}}\overline{Yyx} + \overset{3}{\overline{X}}\overline{zyx} + \overset{4}{ZYXy} + \overset{5}{YXz} + \overset{6}{Xzx} + \quad \text{(first step)}$$

$$+ \overset{7}{\overline{Z}}\overline{Yzy} + \overset{8}{\overline{YX}}\overline{zy} + \overset{9}{\overline{YX}}yz + \overset{10}{ZYzy} + \overset{11}{Zzyx} + \quad \text{(second step)}$$

$$+ \overset{12}{Z}\overline{YXz} + \overset{13}{\overline{ZX}}\overline{zx} + \overset{14}{Yzyx} + \overset{15}{ZYyx} \quad \text{(third step).}$$

We can find other minimal forms. According to Table 1 we can substitute $\bar{Z}YX\bar{y}$ for term 7, $\bar{Y}z\bar{y}x$ for term 9. By deleting term 8 we uncover the vertex A0 only [vertex E0 remains covered by EF01 (10)]. Using the cell AC01 = $= \bar{Z}X\bar{z}\bar{y}$ as substitute for term 8 we cover the vertex A0, spending the same number of operations as before the substitution. The form we get is minimal.

Conclusion

A method has been presented by which it is possible to find at least one minimal form of a given Boolean function. The method uses a map of the function, along with mechanical aids called contact grids and directional grids. The squares of the map carry information about the weights of the corresponding vertices of the body of the given Boolean function.

At least one minimal form is then found in three steps:

1) Terms which must be included in each minimal form are compiled into a partial form.

2) To a partial form which possesses the property of being a part of a minimal form, terms are added in such a way that the said property remains unchanged.

3) From all possible coverings of a given vertex one must belong to a minimal form. A trial is made for each covering under the supposition that 2) is applicable. By comparison of all trials the minimal form is selected.

Other minimal forms are assembled as variations and transformations of the minimal form which has been found. It is interesting to remark that there are cases where some of the steps of the method do not appear.

LITERATURE CITED

1. R. H. Urbano and R. K. Mueller, "A topological method for the determination of the minimal forms of a Boolean function," IRE Trans. EC-5, 126-132 (1956).
2. A. Svoboda, "Graphical-mechanical aids for the synthesis of relay circuits," Nachrichtentechnische Fachberichte, Beihefte der NTZ (Vieweg, Brunswick, 1956), Vol. 4, pp. 213-218.
3. A. Svoboda, "Graficko mechanické pomucky, užívané při analyse kontaktových obvodu [Graphical-mechanical aids used for the analysis of relay circuits]," Information Processing Machines 4, 9-22 (Academy of Sciences, Prague, 1956).

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

ON THE THEORY OF SWITCHING NETWORK SYNTHESIS

F. Svoboda (Prague)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 1071-1079, August, 1961
Original article submitted February 2, 1961

An experimental method is proposed for the synthesis of optimal relay circuits which avoids difficulties connected with the nonexistence of a nontrivial algorithm for the design of such circuits. The method envisages the performance of a series of experiments, each of which leads to obtaining a certain solution (circuit). An experiment consists of a definite number of steps, corresponding to operations of the combinatorial synthesis method.

An optimal alternative at each step is defined as the result of a series of experiments.

1. Introduction

The problem of the optimum. Present experience shows the impossibility of a definition of the optimum suitable for use in all practical cases. Let us illustrate this statement by a simple example.

Let it be required to construct a combinatorial contact network transforming the input variables $i = \{i_1, i_2, i_3, i_4\}$ into the output variables $I = \{I_1, I_2\}$ as given by the table of Fig. 1. The proposed circuit should have one of the following properties:

a) have a minimum number of contacts;
b) have a minimum number of contact springs;
c) have a minimum number of intermediate nodes;
d) have a minimum number of contacts and operate correctly in the presence of incorrect operation* of one (or several) contacts.

i	I
i_1	I_1
i_2	I_2
i_3	I_2
i_4	I_1

Fig. 1.

x_1	x_2	X
0	0	1
0	1	0
1	0	0
1	1	1

Fig. 2. Table of obligatory operating conditions of the circuit.

The conditions of the table (Fig. 1) are obligatory in the sense that any proposed circuit must satisfy these conditions. Conditions a)-d) are optional and may be satisfied for various types of circuits in varying degrees. In the ideal case the form of circuit obtained may satisfy several optional conditions simultaneously. These are examples of differing definitions of the economy or optimalness of circuits.

In the problem considered the carrier of input information can be two binary receiving elements with x_1, x_2 and the carrier of output information-a single two-position active element X. In this case the table of Fig. 1 goes over into the table of Fig. 2. Optimal solutions for conditions a), b), c), are represented in Fig. 3. Evidently, none of these solutions may be considered optimal in the sense of property d). Consider, for example, the second solution of form b), and let us suppose that $x_1 x_2 = 1$. The circuit between the terminals U_0 and X will be closed. If contact x_1 operates incorrectly, the circuit will be open. Thus, the set of optimal solutions corresponding to the differing definitions of optimum need not even intersect.

The choice of optimum condition depends to a substantial degree on the purpose for which the proposed circuit is intended. In principle, there does not exist any definition of optimum which could be usefully employed for all cases met in practice.

It would be desirable to have an algorithm for the synthesis of switching networks giving at least one optimal solution for each optimum condition. For certain simplest conditions of optimum such algorithms are known. An

*Incorrect operation of the contact x_i consists in its being open when with correct operation it should be closed, and vice versa.

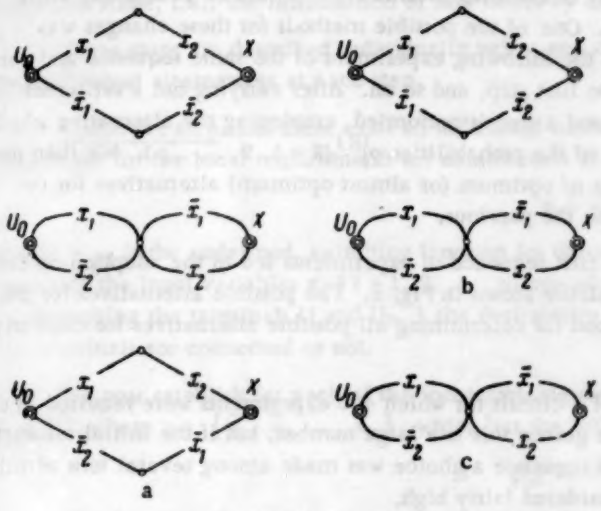


Fig. 3.

example is the algorithm for obtaining the minimal form of a Boolean function given in the disjunctive normal form [1].

However, there are very many optimum conditions and the majority of them are quite complicated, not to mention the fact that more and more conditions are arising in practice. The significance and desirability of studying different optimum conditions in detail in the design of relay systems was pointed out by M. A. Gavrilov.

In the majority of cases an algorithm for the synthesis of optimal relay circuits is unknown, and frequently there is lack of time to find it. The examination of all possible solutions (solution by "exhaustion") is impractical; therefore, it is necessary to proceed to synthesize the circuits by approximate methods. The significance of an experimental approach to the solution of problems in discrete variables was pointed out by A. Svoboda.

The almost universal process discussed in the sequal is applicable to quite varied optimum conditions. However, the realization of this process in practice requires the use of a highspeed computer.

Control by optimum. Let us now described the character of the process and the performance of the experiment. Let a certain optimum condition be given. Starting with the most pessimistic position, let us assume that we have no experience in the design of optimal solutions corresponding to the given condition. Therefore in the course of the first experiment randomly chosen circuits are very uneconomical. Then the frequency of appearance of more economical circuits progressively increases. Each time that a certain "statistical number" of experiments has been carried out, one step• is taken in the direction of the final circuit. The initial data in the method are very simple: it is known that there exists a large number of different solutions, it being assumed that there is a means of obtaining these solutions and that it is possible to estimate the degree of optimalness for each of them, using a certain weighting function. Therefore, when it is necessary to realize a new optimum condition no information on the form of optimal (or almost) circuits is necessary.

The circuit synthesized in Sect. 3 is very simple and it was therefore possible to carry out the process without a highspeed computer. Despite its simplicity and the possibility of obtaining optimal solutions by other methods, the example shows that the process can be used for more complicated cases where standard known methods are inapplicable.

As already mentioned, the basic process is the random determination of circuits satisfying the obligatory conditions of the problem. The final circuit is established progressively, step by step, by the method described by the author in [2]. Before each step there is a sequence of experiments or tests of various alternatives. After a sufficient number of experiments has been carried out, the most advantageous alternative is used to realize the step. The step is carried out by adding to the circuit one or several contacts. Then the process is repeated until the full circuit is obtained.

During a given jth experiment of each sequence of experiments a predefined number of steps is carried out. The choice of alternatives (among ρ alternatives in all) at each step is carried out at random. At the first step of the experiment the alternatives are taken with probabilities $p_j^{(R)}$ ($R = 1, 2, \ldots, \rho$), where for the first experiment of each sequence of experiments $p_1^{(R)} = 1/\rho$. Beginning with the second step the probability of choosing any available alternative is the same and equal to $1/\rho$. The circuit thus obtained after a perfectly defined number of steps, of course, need not give the final solution of the problem, but permits determining the value of the so-called weighting function $f(a^{(j)})$••.

If it is found that $f(a^{(j)}) > \overline{f}_j$, $\overline{f}_j = \sum_1^j f(a^{(j)}) / j$, for the following experiment is is necessary to increase the probability of choosing the alternative R employed in the first step of the preceding experiment and to decrease the probabilities of choosing the remaining alternatives $1, 2, \ldots, R-1, R+1, \ldots, \rho$, in proportion to their values

• Translator's note: The author uses the word "step" here to mean the introduction of a branch or a new node in an as yet incomplete network. See below, Section 2.

•• The method for determining the weighting function is described below.

in the preceding experiment. If $f(a^{(j)}) \leqslant \bar{f}_j$, the probability of choosing alternative R should be decreased, while the remaining alternatives increased (or remain unchanged). One of the possible methods for these changes was described in [3]. Carrying out these changes, we proceed to the following experiment of the same sequence and correct the probabilities for random choice of alternatives of the first step, and so on. After carrying out a sufficient number of experiments the process is considered completed and a step is performed, employing the alternative which corresponds to the maximum or one of the maximum values of the probabilities $p_j^{(R)}$ ($R = 1, 2, \ldots, \rho$). We then proceed to the following sequence of experiments for the choice of optimum (or almost optimum) alternatives for the next step of the synthesis. This sequence is exactly similar to the previous.

With regard to the example considered in Sect. 3, the first sequence of experiments led to the adoption on the first step of the alternative $\bar{x}_1 + x_2$ ($\alpha^{(8)}$) of the nine possibilities shown in Fig. 8. The possible alternatives for the second series of experiments are given in Fig. 10. The method for determining all possible alternatives for each step is described in Sect. 2.

Sect. 3 describes in condensed form the construction of a circuit for which 400 experiments were required in the first series and 100 experiments in the second series. At first glance this is a large number, but if the initial assumptions are taken into account, together with the fact that in the first sequence a choice was made among several tens of millions of possible circuits, the efficiency of the process may be considered fairly high.

In contrast to the weighting function for constructing the Posthumus sequence [3], in the weighting function employed here there is no simple method for finding in advance either the number of variables or the region in which these variables is defined. We shall clarify this by Fig. 8. In this figure all possible alternatives for the first two synthesis steps of the illustrative circuit are given. To each path, beginning at point s and terminating at the point marked by double circles, there corresponds a definite form of circuit, and a definite integer value of the weighting function f (a) depends on the alternatives of all the steps, i.e., $a = \{a_0, a_1, \ldots, a_\mu, \ldots, a_{\mu k}\}$, where a_μ is the set of alternatives of the first step. In Fig. 8, for the first two steps we obtain:

$$a_0 = \varnothing\{, x_1, \bar{x}_1, \bar{x}_1 + x_2, \ldots, x_1 + \bar{x}_2, \bar{x}_1 + \bar{x}_2\},$$

$$a_1 = \{\varnothing; \bar{x}_1, \bar{x}_1 + x_2, x_2, \varnothing; x_1, x_1 + \bar{x}_2, \bar{x}_2, \varnothing;$$

$$\ldots; \bar{x}_1, \bar{x}_1 + x_2, x_2, \varnothing; \varnothing\}.$$

From what has already been said and from the fact that there exists a tremendous number of different circuits, it may be concluded that the number of variables cannot be determined in a simple manner; in exactly the same way it is practically impossible to determine the region of these variables, beginning already from $a_3, \ldots$. The range of values of the weighting function f (a) is also unknown, it being known only that the function takes on an integer value (in our case the value of the function corresponds to the number of intermediate nodes). However, if we substitute the first (μ -1) variables and proceed to the experimental sequence of the next step, it is always possible to define the region of variation of the variable a_μ. Furthermore, the value of the function f (a) can be found even when the values of the substituted variables do not correspond to the final circuit. Thus, in our case even for an unfinished circuit, it is possible to determine the number of intermediate nodes and, in this way, to find the value of the function.

The possibility of an experiment to obtain the optimal solution of the problem evidently depends on the circumstance that if the weighting function for a given substitution of the variable (or one of the alternatives) takes on an extremal value, it will take on for this same substitution a sufficient number of other values not differing substantially from the extremum.

2. Determining the Alternatives

We shall use the combinatorial method [2] for circuit design. We described here a somewhat altered and improved form of the method. A synthesis step corresponds either to the operation ω_1 (inductive step) or the operation ω_2 (deductive step). The designations of the steps are introduced in accordance with the concepts of the theory of cognition. By the inductive method we understand the introduction into consideration of new, as yet unknown elements; by the deductive method – the search for new relationships or dependencies among already known facts or mathematical objects. The deductive stage of the method is exhausted by the use of the operation ω_2. If after this it is necessary to continue the solution (if the synthesis of the circuit is not terminated) it is necessary to pass to the

inductive stage, i.e., the introduction of new nodes by the operations ω_1, and then again to pass to deductive steps.

These steps are described individually below and certain terminology is introduced to give a general concept of the individual alternatives at each step.

Deductive step. Let there exist an unfinished circuit, being designed by the combinatorial method. We denote the matrix for the local requirements for construction of the circuit

$$\Psi = \|\Psi_{gh}\|,$$

where Ψ_{gh} is the undefined switching function for the two nodes U_g and U_h, which takes on for each combination of values of the input variables x_i ($i = 1, 2, \ldots, n$) one of the three values 0, 1, or ~; here 0 indicates the impossibility of connecting the terminals U and U_n, 1 the desirability of connecting the terminals, and ~, that it is indifferent whether the terminals are connected or not.

We now establish for each of the indeterminate switching functions Ψ_{gh} the maximum function of the Σ-branch $f_{gh}(x)$, where $x = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\}$, which has the following property:

$$\text{if} \quad \psi_{gh} = 0, \text{ then } f_{gh}(x) \neq 1. \tag{2.1}$$

The function $f_{gh}(x)$ is greatest in the sense that each function having the property (2.1) is already included in it.

Let us put

$$f_{gh}(x) = \tilde{x}_{\nu_1}^{gh} + \tilde{x}_{\nu_2}^{gh} + \ldots + \tilde{x}_{\nu_{gh}}^{gh},$$

where $\nu_1, \nu_2, \ldots, \nu_{gh} \in N$, $N = 1, 2, \ldots, n$; $\tilde{x} = x$ or $\tilde{x} = \bar{x}$.

The switching elements in all Σ-branches obtained in this matter will be denoted for conciseness by $p_1, p_2, \ldots, p_i, \ldots, p_\nu$, where p_i are variables denoting ϕ in the case where the element is not used in the deductive step and $\tilde{x}_{\nu_{gh}}^{gh}$ in the case where the element is used. With ν variables there are 2^ν possibilities for selecting their values and the same number of possible alternatives. We thus find the basic system of alternatives for the deductive stage.

We now classify the alternatives among permissible and impermissible. An impermissible alternative, in contrast to a permissible one, cannot be realized, since it gives rise to a sneak path (violation of the obligatory conditions of the proposed circuit).

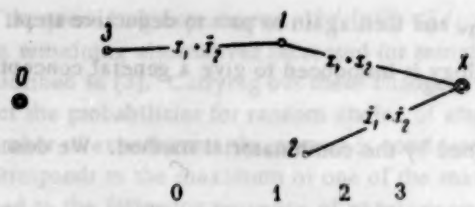
Let us summarize the above. A deductive step is carried out, on the basis of the matrix $\|\Psi_{gh}\|$, of local requirements for the proposed circuit. For each element of the matrix Ψ_{gh} we find the corresponding maximum Σ-branch. The set of all possible Σ-branches of length ν defines the maximum possible alternative. The 2^ν subsets of this alternative define all possible alternatives.

As an example let us consider experiment 377 of the first sequence of experiments. Fig. 4 shows the partial circuit, its matrix of locally possible Σ-branches, and the table of basic systems of alternatives. In the matrix the elements x_i and $\bar{x}_i$ are chosen randomly from the set of all possible Σ-branches, since for them the undefined function Ψ_{gh} permits all Σ-branches.

Inductive step. We again assume an unfinished circuit, and that continuation of the solution by a deductive step is undesirable or impossible.

We denote the matrix of local requirements for constructing the circuit at the inductive step by $\Phi = \|\varphi_k\|$ where φ_k is the undefined switching function of the node U_k, which for each combination of values of the input variables x_i ($i = 1, 2, \ldots, n$) takes one of the following values: 0, 1, or ~. These values have the following significance: 0 — desirable to obtain $f_k(x) = 0$; 1 — desirable to obtain $f_k(x) = 1$; ~ indicates that it is indifferent which value the function $f_k(x)$ takes.

In the table of Fig. 5 the variable p_i takes on the value ϕ or $\tilde{x}_i$ or $\bar{\tilde{x}}_i$. For each of the functions φ_k it is possible to use one of the 3^n alternatives. In all there are 3^{Kn} alternatives for the K nodes of the circuit. They are all permissible. As a rule, in the inductive step there is a much larger number of alternatives than in the deductive step, since they are not restricted by the obligatory conditions of the proposed circuit.



	0	1	2	3
0	1			
1	/	1		
2	/	x_1	1	
3	$\bar{x}_1 + x_2$	/	$\bar{x}_1$	1

a

p_1	p_2	p_3	p_4	Admissibility
Ø	Ø	Ø	Ø	yes
Ø	Ø	Ø	$\bar{x}_2$	yes
Ø	Ø	x_2	$\bar{x}_2$	no
Ø	Ø	x_2	Ø	yes
Ø	$\bar{x}_1$	x_2	Ø	yes
Ø	$\bar{x}_1$	x_2	$\bar{x}_2$	no
Ø	$\bar{x}_1$	Ø	$\bar{x}_2$	yes
Ø	$\bar{x}_1$	Ø	Ø	yes
x_1	$\bar{x}_2$	Ø	Ø	no
x_1	$\bar{x}_1$	Ø	$\bar{x}_2$	no
x_1	$\bar{x}_1$	x_2	$\bar{x}_2$	no
x_1	$\bar{x}_1$	x_2	Ø	yes
x_1	Ø	x_2	Ø	yes
x_1	Ø	x_2	$\bar{x}_2$	no
x_1	Ø	Ø	$\bar{x}_2$	yes
x_1	Ø	Ø	Ø	yes

b

Fig. 4. Example of basic system of alternatives for deductive step: a) matrix of locally possible Σ-branches; b) table of basic systems of alternatives.

p_1	. . .	p_i	. . .	p_{n-1}	p_n
Ø		Ø		Ø	Ø
Ø		Ø		Ø	x_n
Ø		Ø		Ø	$\bar{x}_n$
Ø		Ø		x_{n-1}	Ø
Ø		Ø		x_{n-1}	x_n
Ø		Ø		x_{n-1}	$\bar{x}_n$
Ø		Ø		$\bar{x}_{n-1}$	Ø
Ø		Ø		$\bar{x}_{n-1}$	x_n
.		.	.	.	.
.		.	.	.	.
.		.	.	.	.
$\bar{x}_1$	. . .	$\bar{x}_i$	. . .	$\bar{x}_{n-1}$	$\bar{x}_n$

Fig. 5. Table of system of alternatives for operation ω_1.

3. Experiment

It is necessary to construct a combinatorial contact network which satisfies the obligatory conditions given by the table of Fig. 6 and has a minimum number of nodes.

A series of experiments will be performed for the design of the network. It is first necessary to decide with which step—deductive or inductive—the synthesis begins. There is as yet no definite criterion for the choice of type of step. It should be kept in mind that an inductive step is always possible, while a deductive step, only in certain cases. In the general case, when either a deductive or an inductive step is possible, the choice of type of step (similarly, the choice of alternatives in these steps) is preferably carried out on the basis of the experimental sequence. In our case the first step can only be an inductive step.

In the initial circuit (Fig. 7a) there is only a single node, permitting expansion of the function X. In Fig. 8 nine possible alternatives for this step are shown. For the choice of the best, (or nearly best) alternative, a sequence of 400 experiments was carried out. To each of these nine alternatives there corresponds a controlling random function [3]. The functions are represented in Fig. 9. The greater the value taken on by the controlling random function the more advantageous is the corresponding alternative. Finally, alternative $\alpha^{(8)}$ was selected at random from the alternatives $\alpha^{(8)}$ and $\alpha^{(7)}$. The void alternative cannot be used, since this implies repetition of the sequence of experiments.

x_1	x_2	X
0	0	1
0	1	0
1	0	0
1	1	1

Fig. 6.

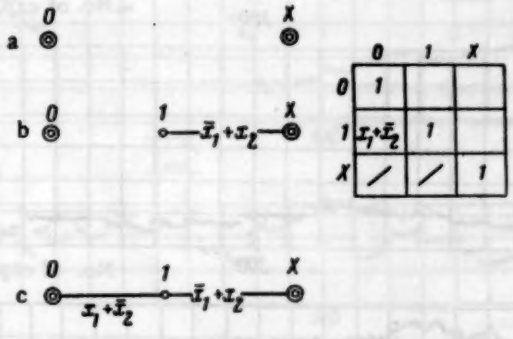


Fig. 7. Stages of solving the problem: a) initial state; b) after first inductive step; c) after first deductive step.

1. Analyzing the problem, we come to the conclusion that all the alternatives can be divided into four classes of equivalent alternatives: $\alpha^{(3)}, \alpha^{(7)}; \alpha^{(1)}, \alpha^{(2)}, \alpha^{(5)}, \alpha^{(6)}; \alpha^{(4)}; \alpha^{(8)}; \alpha^{(0)}$. We take equivalence in the sense that using alternatives of a given class leads to circuits for which the values of the weighting functions are almost identical. Four hundred experiments were required for the experimental establishment of this fact. It should be kept in mind that 100 experiments are fully satisfactory for a qualitative choice of alternatives.

2. The logical feedback is based on the following principle: If it is found that any given alternative leads systematically to a better result, the probability (and thus its frequency) of its choice is increased. As a result of this the corresponding control random function takes on a stable value sooner than the other functions. The remaining functions are delayed in the sense that the stabilization of their values occurs later. This process has the following essential property. Let us assume that a disadvantageous alternative was found satisfactorily in several experiments. Although the probability of choice of this alternative increases during these experiments, at first disadvantageous result it drops sharply. The process proceeds similarly for the advantageous alternatives.

Deductive step. The matrix of local requirements is given in Fig. 7b. There are four alternatives available (Fig.10).

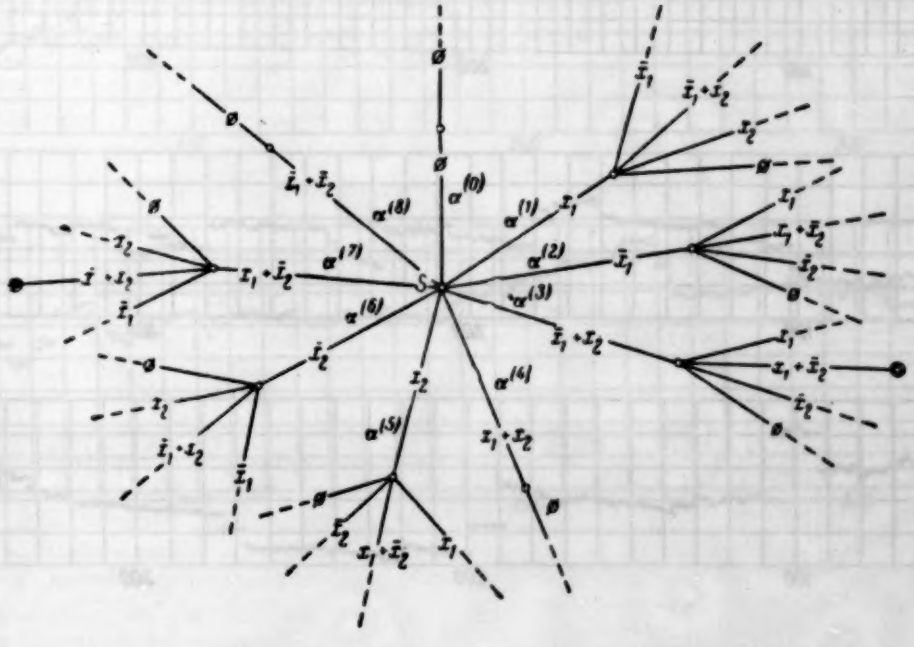
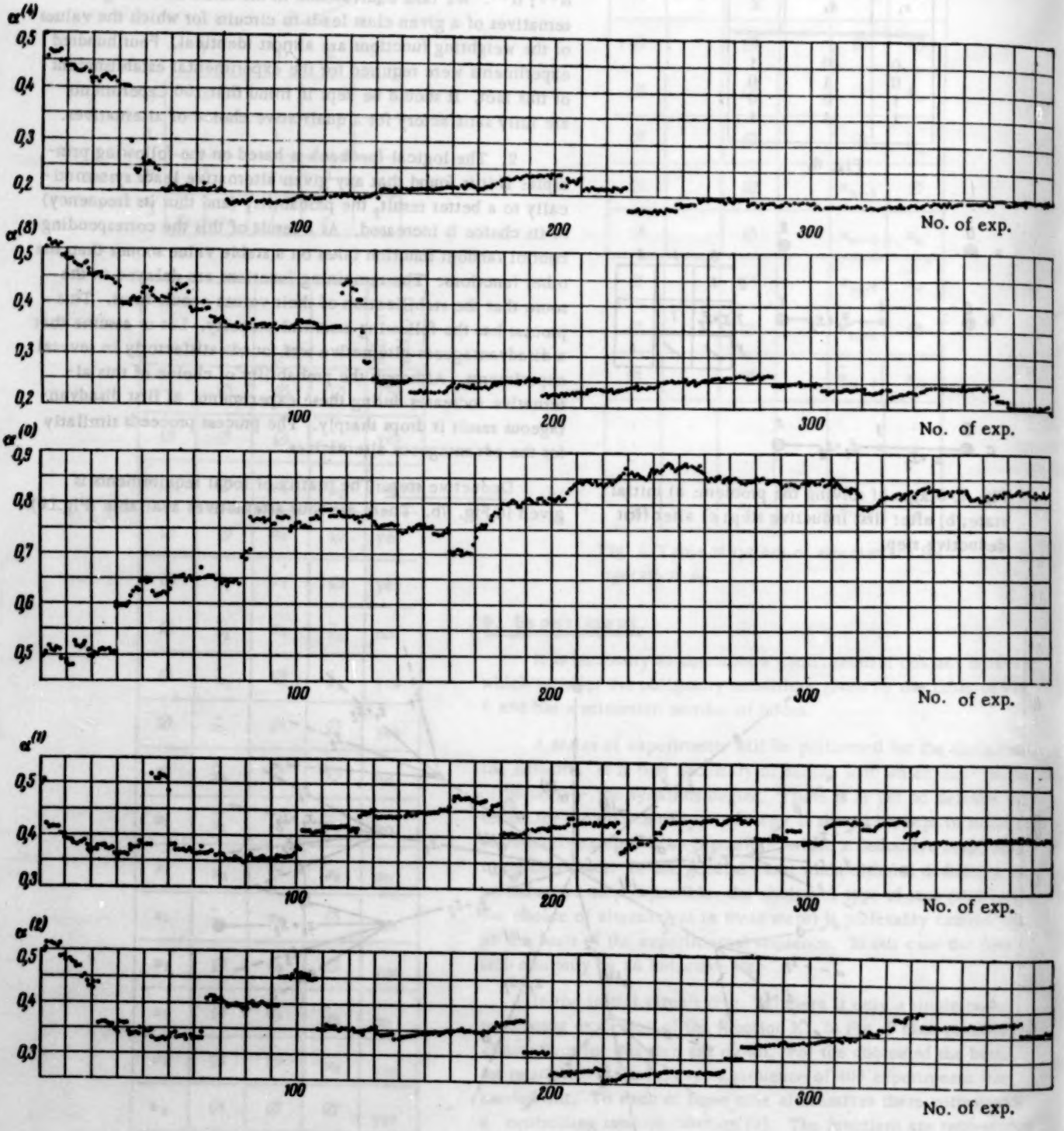
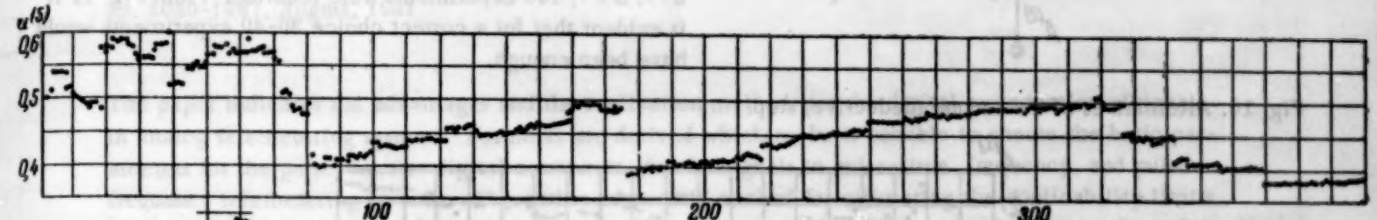
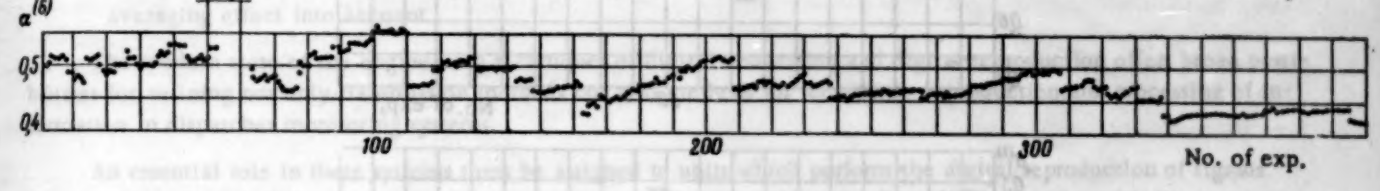
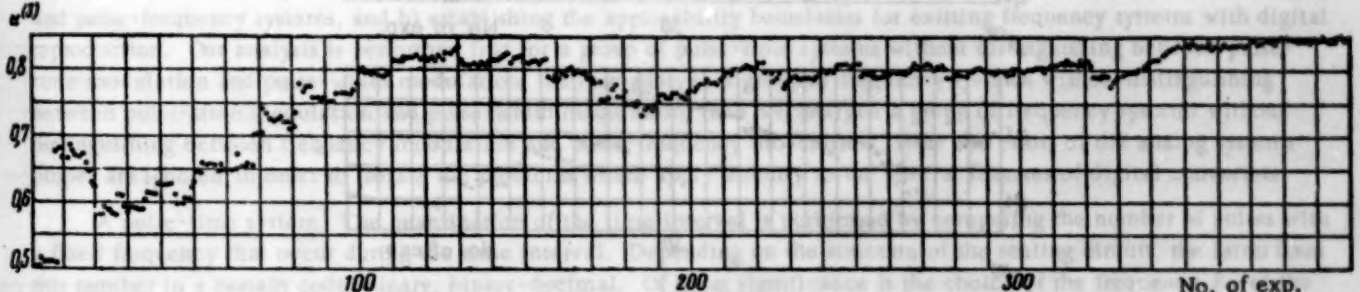


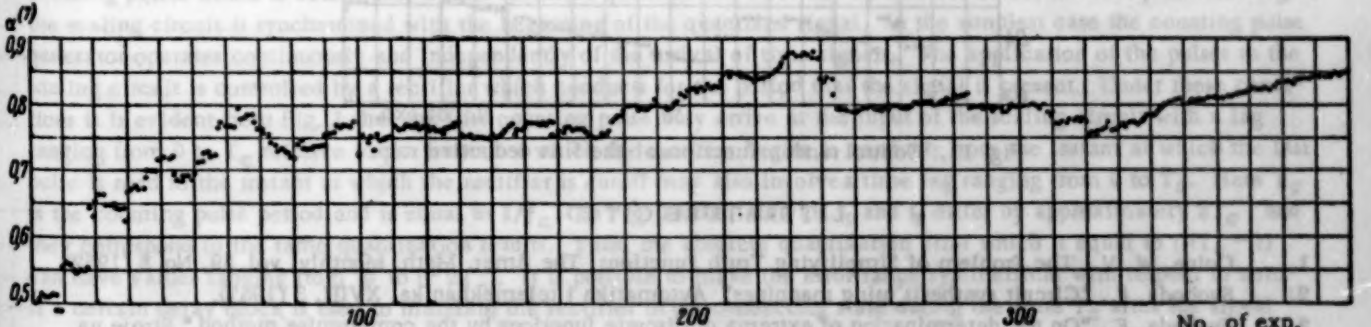
Fig. 8. Alternatives of the first two steps of solving the problem.


α(4)
0,5
0,4
0,3
0,2
100
200
300
No. of exp.
α(8)
0,5
0,4
0,3
0,2
100
200
300
No. of exp.
α(0)
0,9
0,8
0,7
0,6
0,5
100
200
300
No. of exp.
α(1)
0,5
0,4
0,3
100
200
300
No. of exp.
α(2)
0,5
0,4
0,3
100
200
300
No. of exp.


α(5)
0,6
0,5
0,4
100
200
300
No. of exp.


α(6)
0,5
0,4
100
200
300
No. of exp.


α(3)
0,8
0,7
0,6
0,5
100
200
300
No. of exp.


α(7)
0,9
0,8
0,7
0,6
0,5
100
200
300
No. of exp.

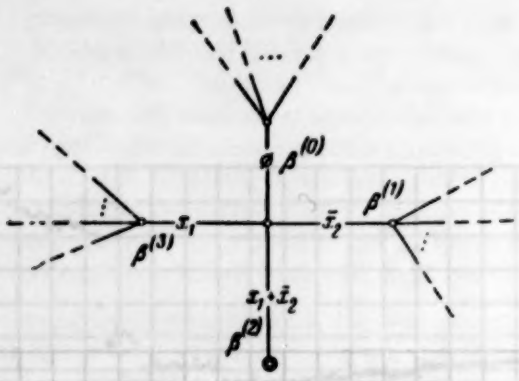


Fig. 10. Alternatives of the second (deductive) step.

For choice of the best alternative a sequence of 100 experiments was carried out; the values of the corresponding control random functions are given in Fig. 11. The experiment led to the final choice of alternative $\beta^{(2)}$. The result obtained, shown in Fig. 7c, actually corresponds to the optimal solution.

In this sequence of experiments there are three classes of equivalent alternatives: $\beta^{(2)}$; $(\beta)^{(1)}$, $\beta^{(3)}$; $\beta^{(0)}$. For the experimental verification of the equivalence of alternatives $\beta^{(1)}$, $\beta^{(3)}$, 100 experiments were required. From Fig. 11 it is evident that for a correct choice 30-40 experiments would have been enough.

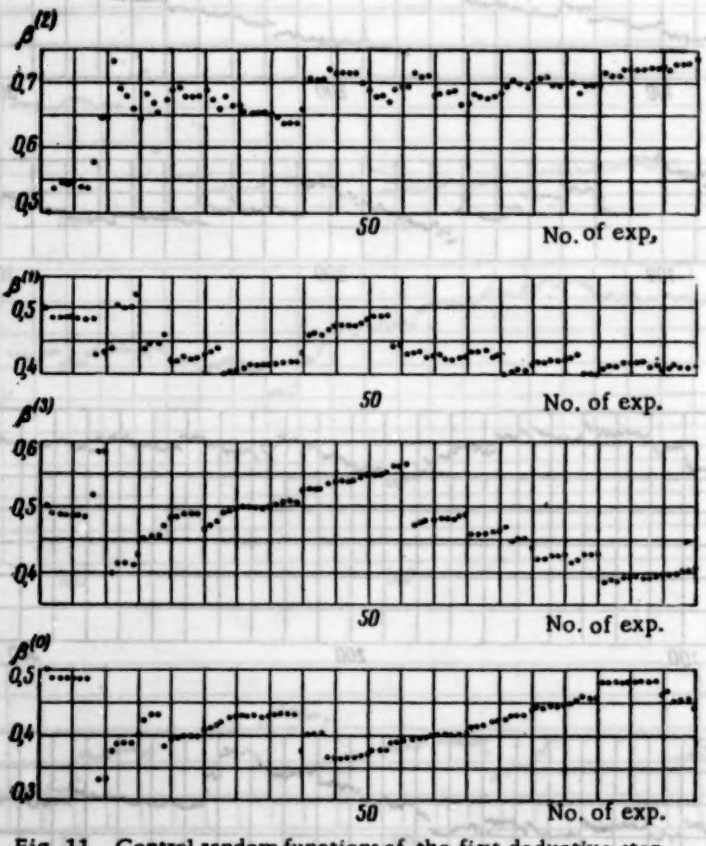


Fig. 11. Control random functions of the first deductive step.

LITERATURE CITED

1. Quine, W. V., The Problem of Simplifying Truth Functions, The Amer. Math. Monthly, vol. 59, No. 8 (1952).
2. Svoboda, F., "Circuit synthesis using machines," Avtomatika i telemekhanika, XVIII, 3 (1957).
3. Svoboda, F., "On the determination of extrema of discrete functions by the compromise method," Stroje na zpracovani informaci, V (1957).
4. Bibliography of domestic and foreign literature for 1950-1954, Avtomatika i telemekhanika, XVI, 4, 409-420 (1955).

ON DIGITAL REPRODUCTION OF SIGNALS IN ANALOG TELEMETERING SYSTEMS

Ya. A. Kupershmidt (Moscow)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 1080-1087, August, 1961

The paper indicates the advantages and the realization methods for the digital reproduction of signals in analog telemetering systems. Formulas are derived which make it possible to choose the basic parameters for the pulse-counter digital converters for the signals in pulse-time, frequency, and pulse-frequency telemetering systems. The author proposes a method for estimating the applicability limits of the digital reproduction of signals in existing frequency systems; the method is based on taking the averaging effect into account.

The modern state of the engineering techniques of digital conversion and digital reproduction offers broad possibilities for refining not only transmission methods but also methods for reception, reproduction, and processing of information in dispatcher monitoring systems.

An essential role in these systems must be assigned to units which perform the digital reproduction of signals from analog telemetering systems. In this paper we are evidently making the first attempt at a) establishing certain quantitative relationships for choosing the basic parameters in digital converters for the signals of pulse-time, frequency, and pulse-frequency systems, and b) establishing the applicability boundaries for existing frequency systems with digital reproduction. Our analysis is performed first for a group of pulse-time systems without distinguishing between pulse-time modulation and pulse-width modulation; then we analyze a group of frequency systems without distinguishing between pulse-time modulation and pulse-width modulation; then we analyze a group of frequency systems without distinguishing between frequency modulation and pulse-frequency modulation. Here the errors of the analog systems proper are ignored in order to isolate the problems which apply directly to the special features of digital converters.

A pulse-time system. The quantization of the time interval is performed by computing the number of pulses with a fixed frequency that occur during the time interval. Depending on the structure of the scaling circuit, the latter fixes this number in a certain code: binary, binary-decimal. Of great significance is the choice of the frequency f_c of the counting pulses which is determined by the specified quantization error and by whether or not the first pulse arriving at the scaling circuit is synchronized with the beginning of the quantized signal. In the simplest case the counting pulse generator operates continuously and independently of the arrival of time signals. The application of the pulses to the scaling circuit is controlled by a rectifier which conducts for the period that the signal is present. Under these conditions it is evident from Fig. 1 that the first counting pulse may arrive at the input of the scaling circuit with a lag ranging from 0 to T_c relative to the instant at which the rectifier begins to conduct; from the instant at which the last pulse is read to the instant at which the rectifier is cutoff may also involvea time lag ranging from 0 to T_c. Here T_c is the counting pulse period and is equal to $1/f_c$. In Fig. 1a the intervals t_1 and t_2 differ by approximately $2T_c$, and they correspond to the same quantization results. Thus, the absolute quantization error which is equal to $(NT_c - t)$ can have values ranging from 0 to $(-2T_c)$. It is possible to make the error range symmetrical with respect to zero if a certain delay block is used to maintain the rectifier in the conducting state during the time T_c after the signal has been completed. This function can be performed by a one-shot multivibrator with a delay time T_c which is triggered by the trailing edge of the coded pulse. Then the absolute quantization error acquires values from $-T_c$ to $+T_c$.

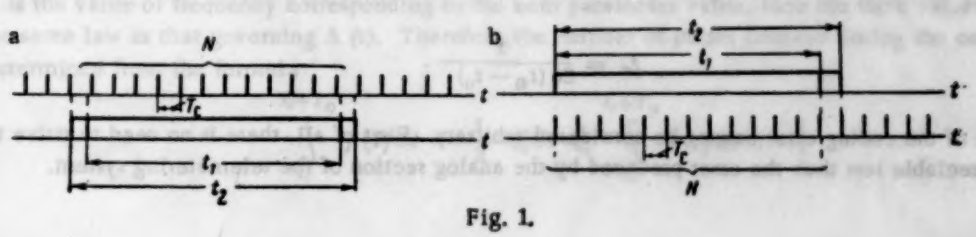


Fig. 1.

If the counting pulse generator must be triggered by the leading edge of the signal, then it is evident from Fig. 1b that the error acquires values from 0 to $-T_c$; when a delay block is used which produces a delay equal to $T_c/2$, the error acquires values from $-T_c/2$ to $+T_c/2$. Usually in pulse-time telemetering systems the zero parameter value is transmitted by a pulse with a certain duration t_0, so that a variation of the parameter A from 0 to the nominal A_n corresponds to the variation of the pulse duration from t_0 to t_n. Between A and $\underline{t}$ there is the following relationship:

$$t = t_0 + \frac{t_n - t_0}{A_n} A.$$

Therefore, in coding it is necessary to subtract a certain number N corresponding to the total value of $\underline{t}$ from the number N_0; the value N_0 corresponds to the value t_0. This subtraction is achieved quite simply by setting the code for the number $(N_{lim} - N_0 + 1)$ in the scaling circuit; here N_{lim} is the limiting number which can be fixed by the counter. This method of presetting the initial state of the counter is also useful for achieving a symmetrical quantization error in the case where the peak-to-peak fluctuations of this error are equal to an even number of intervals T_c. If the error varies over the range 0 to $2T_c$, then in order to make it symmetrical it is sufficient to introduce an additional "one" into the counter before starting the count after setting the number $(N_{lim} - N_0 + 2)$.

The number N_0 proper can represent the initial interval with an error ranging from $-T_c/2$ to $+T_c/2$ (i.e., the difference $N_0 T_c - t_0$ may fluctuate within these limits). As a result of subtracting N_0 from the number N the errors can be added with identical signs. The error involved in quantizing the interval t_0 can be eliminated by designing a telemetering system with the value of t_0 selected so that an integral number of counting frequency f_c periods can be fitted into it. It is possible to achieve a less rational solution; namely, to add a delay block (a time relay) at the receiving end which begins counting after a time t_0 has elapsed, measured from the leading edge of the time pulse arriving from the communication channel.

The normalized coding error δ is equal to

$$\delta = \frac{(N - N_0) T_c - (t - t_0)}{t_n - t_0}.$$

If we originate from the specified normalized quantization error $\pm\delta$, then we can find the minimum allowable counting pulse frequency f_c for different variants of the counting circuit described above. Thus, for the case where the counting pulse generator is not synchronized with the signal and the initial interval t_0 is quantized with an error $\pm T_c/2$, the total quantization error reaches $\pm 3T_c/2$ (if there is a delay block that produces a symmetrical error for quantization of the interval $\underline{t}$, or if the process of achieving symmetry is performed by introducing an additional unit when the initial setting of the counter is made).

Thus

$$\pm\delta = \frac{\pm 3T_c}{2(t_n - t_0)}.$$

Taking into the account the fact that $T_c = 1/f_c$, we find

$$f_{c1} = \frac{3}{2\delta(t_n - t_0)}.$$

Note that for such a counting method the lowest order place in the code becomes unreliable, since the error exceeds the value of that place. In analogous fashion we compute the required counting frequency for the case where there is no error involved in quantizing t_0, the counting pulse generator is synchronized with the signal, and the delay block produces a quantization error which is symmetrical and confined within the limits $\pm T_c/2$. Under these conditions

$$f_c = \frac{1}{2\delta(t_n - t_0)}.$$

The choice of the coding error cannot be considered arbitrary. First of all, there is no need to strive to make that quantity appreciable less than the error produced by the analog section of the telemetering system.

But this concept is not the only one. The choice of the amplitude subdivision steps is closely associated with the choice of the time subdivision steps. For analog reproduction the time subdivision steps are determined by the upper boundary of the frequency spectrum for the information function and can be computed from the well known Kotel'nikov theorem. The use of a low-pass filter at the system output makes it possible to reproduce the form of the information function in the intervals between separate measurements.

For the case of digital reproduction there is no such filter. In the intervals between measurements a constant digital reading is maintained. The use of such readings is possible only for the condition that for the fastest variation of the function two adjacent readings will differ only by an amount equal to one unit in the lowest order place (i.e., by one subdivision step). An exception can be made for cases of emergency step variation of the measured parameter. Such a requirement makes the choice of the amplitude subdivision step depend on the time subdivision step. If digital reproduction is adapted to an existing analog telemetering system for which the difference $(t_n - t_0)$ and the time interval between individual measurements of the ordinates T for the specified function are known, then the value of δ should be chosen so that for the maximum rate of change of the parameter A in a normal working mode

$$a = \left| \frac{dA}{dt} \right|_{\max}$$

the value of A has time to vary by no more than $A_n\delta$, where A_n is the scale nominal. Note that the quantity T is proportional to the number of telemetering channels that are serviced by one digital converter at the dispatcher point. The modulus of the increment of the function during the period T is equal to aT. In accordance with the conditions cited above

$$aT \leqslant A_n\delta.$$

From this we have

$$\delta \geqslant \frac{aT}{A}.$$

If the value of δ computed in this fashion is too large, then digital reproduction should be rejected.

Frequency and pulse-frequency systems. Digital conversion of the signals in these systems is also performed using a pulse counter circuit. The difference between such a circuit and time signal converters consists of the fact that the stable counting pulse generator is replaced by a pulse shaper that is controlled by a signal arriving from the channel; the instant, at which the rectifier at the input of the scaling circuit becomes conducting, controls the stable time interval generator (a stable time relay). Under these conditions the number of counted pulses is proportional to the signal frequency. However, there is a more substantial difference between digital frequency conversion and digital time conversion. It consists of the presence of averaging which also occurs for analog reception of a frequency signal. The averaging effect imposes additional conditions on the selection of the parameters for a digital conversion system.

The time during which each averaging is performed must be chosen on the basis of the dynamics governing the monitored parameter and on the basis of the specified error involved in reproducing that parameter. Here it is necessary to take into account two types of errors: the dynamic error associated with averaging, and the error involved in the quantized nature of the readings. Of special importance is the problem associated with the rates of parameter variation for which digital reception of the signal is usable in some existing frequency telemetering system which was initially designed for analog reproduction of that parameter.

If the frequency f is a linear function of the measured parameter A (t)

$$f(t) = f_0 + \frac{f_n - f_0}{A_n} A(t),$$

where f_0 is the value of frequency corresponding to the zero parameter value, then the time variation of the frequency obeys the same law as that governing A (t). Therefore, the number of pulses counted during the counting interval T_c can be determined from the formula

$$\int_{t_1}^{t_1+T_c} f(t)\,dt - \frac{1}{2} \leqslant N \leqslant \int_{t_1}^{t_1+T_c} f(t)\,dt + \frac{1}{2}.$$

Here t_1 is the time from which the next reading begins. The number N is the only integer which is within the indicated interval.

The average frequency in the interval $(t_1, t_1 + T_c)$ is assumed equal to

$$f_{av} = \frac{N}{T_c}.$$

This quantity is normalized in relative or absolute units of the parameter A and is stored by the digital indicator from the instant $t_1 + T_c$ until a new value f'_{av} is obtained as a result of a new scaling operation. The metering error should be assumed to be the maximum deviation of the frequency f from f_{av} over the interval extending from time t_1 to the instant at which a new value f'_{av} is obtained. For the case of a multichannel digital converter with time division that services several frequency telemetering systems, the counting intervals are not adjacent but are spread apart by the amount $T' = (n - 1) T_c$, where n is the number of channels. For $n = 1$ we have $T' = 0$.

Therefore, the metering error is assumed to be equal to the maximum deviation of the value f_{av}, found at the instant $t_1 + T_c$, from the value of f in the interval from $t_1 + T_c$ to $t_1 + T_c + T' + T_c = t_1 + 2T_c + T'$. We shall assume that the maximum rate of change of the measured parameter A is known; i.e., we know the value of

$$a = \left| \frac{dA}{dt} \right|_{\max}.$$

We shall assume that we are dealing with the worst case, in which the parameter A increases at the maximum rate during the period extending from the instant t_1 to the instant $t_1 + 2T_c + T'$: $dA/dt = a$ in the interval $[t_1, t_1 + 2T_c + T']$. We shall estimate the digital reproduction error which is obtained under these conditions. In the interval under study

$$A(t) = A(t_1) + a(t - t_1).$$

For the sake of brevity we shall use the expression

$$\frac{f_n - f_0}{A_n} = k.$$

Then

$$f(t) = f_0 + kA(t) = f_0 + kA(t_1) + ka(t - t_1).$$

The number of pulses computed in the interval $[t_1, t_1 + T_c]$ is defined according to the formula

$$N = \int_{t_1}^{t_1 + T_c} [f_0 + kA(t_1) + ka(t - t_1)]\, dt - \Delta N.$$

Here ΔN is the quantization error which is determined by the fact that we cannot fit an integral number of signal periods into the interval T_c. We shall discuss the quantity ΔN later. Performing the integration and dividing N by T_c, we obtain the value of f_{av} which sets the digital receiver:

$$f_{av} = f_0 + kA(t_1) + \frac{kaT_c}{2} - \frac{\Delta N}{T_c}.$$

For a linear rise of the signal with the slope the frequency reaches the following value at the instant at which a new digital value is obtained (i.e., at the instant $t_1 + 2T_c + T'$):

$$f(t_1 + 2T_c + T') = f_0 + kA(t_1) + ka(2T_c + T').$$

Therefore, the maximum absolute error of digital reproduction for a rising signal will be equal to

$$\Delta f^{(+)}_{\max} = f_{av} - f(t_1 + 2T_c + T') = -\frac{3}{2} kaT_c - kaT' - \frac{\Delta N}{T_c}.$$

In the case where the measured parameter decreased at the same maximum rate $\underline{a}$ we would have the reproduction error

$$\Delta f^{(-)}_{\max} = +\frac{3}{2}kaT_c + kaT' - \frac{\Delta N}{T_c}.$$

Therefore,

$$\Delta f_{\max} = \pm\left(\frac{3}{2}kaT_c + kaT'\right) - \frac{\Delta N}{T_c}.$$

The normalized relative error is computed from the formula

$$\delta f = \frac{\Delta f_{\max}}{f_n - f_0} = \frac{\Delta f_{\max}}{kA_n} = \pm\frac{a}{A_n}\left(\frac{3}{2}T_c + T'\right) - \frac{\Delta N}{(f_n - f_0)T_c}.$$

Taking into account the fact that $T' = (n-1)T_c$, we obtain

$$\delta f = \pm\frac{aT_c\left(\frac{1}{2}+n\right)}{A_n} \cdot \frac{\Delta N}{(f_n - f_0)T_c}.$$

The first term in this expression denotes a dynamic error that is associated with the delay which exists between the results of the average measurements and the time variations of the parameter. The second term denotes the error in the discrete nature of the measurements of the average frequency over the interval T_c. For the case where the beginning of the reading is synchronized with the leading edge of one of the signal pulses and the error is then made symmetrical for digital reproduction, we can introduce ΔN within the limits $\pm\frac{1}{2}$, as was indicated above. For the case of non-synchronous counting the value of N may be too small by an amount ranging from 0 to 2. Subsequent symmetrization of the error for digital reproduction can be used to introduce ΔN within the limits ± 1. We subtract the number N_0 from the number N, where N_0 corresponds to the initial frequency f_0 with an accuracy of up to $\pm\frac{1}{2}$. Therefore, the total quantization error reaches the value

$$\Delta N = \pm\frac{3}{2}.$$

When the above is taken into account the expression for estimating the normalized digital reproduction error for a frequency signal becomes

$$\delta f = \pm\frac{aT_c\left(\frac{1}{2}+n\right)}{A_n} \pm \frac{3}{2}\frac{1}{(f_n - f_0)T_c}.$$

The expression determined above makes it possible to compute the required counting interval duration and the allowed rate of change of the parameter from the specified values of the quantization errors and the dynamic error. It is assumed that the parameters of the frequency telemetering system are known (i.e., we know f_n and f_0). In specifying the magnitude of the digital reproduction error we should be guided by the accuracy with which the frequency signal isolated by the analog receiver represents the telemetered parameter. There is no sense in requiring the accuracy of digital reproduction to be higher than the accuracy of the analog section of the system. We shall specify identical values of the quantization error and the dynamic error. We shall denote this error by the letter δ. Then we obtain two computing equations:

$$\delta = \frac{3}{2(f_n - f_0)T_c}, \qquad \delta = \frac{aT_c\left(\frac{1}{2}+n\right)}{A_n}.$$

From the first equation we compute the magnitude of the counting interval T_c:

$$T_c = \frac{3}{2(f_n - f_0)\delta}.$$

Substituting this expression into the second equation and solving the latter with respect to $\underline{a}$, we obtain

$$a = \frac{2\delta^2 A_n(f_n - f_0)}{3\left(\frac{1}{2}+n\right)}.$$

The latter equation makes it possible to establish the limiting allowable rate of change of the parameter in the case of centralized digital reception of the signals from n frequency telemetering systems with known values of the minimum and maximum frequencies. Note that there is a quadratic relationship between the allowable rate of change a of the parameter and the magnitude of the error δ. This means that the reduction of the error by a factor of m is achieved at the expense of reducing the maximum rate of change of the reproduced parameter by a factor of m^2.

It is natural that the problem arises as to what relationship there is between the allowable rates of variation for the parameter in the two cases where one and the same frequency (or pulse-frequency) telemetering system is used with 1) analog reproduction, and 2) digital reproduction. It is assumed that the comparison is performed under identical conditions with respect to number of channels (i.e., for single-channel digital reproduction for which n = 1). In order to perform this estimate we find the expression for determining the allowable rate of rise of the parameter for analog reproduction of the frequency signal. As we noted above, both analog reproduction and digital reproduction are associated with mandatory time averaging. This is determined by the necessity of eliminating vibrations of the pointer on the instrument that measures the average current produced by the pulses that carry charge dosages.

Based on the above, we can represent an analog frequency receiver in the form of a generalized circuit containing an inertialess section which performs linear conversion of the frequency f into the voltage U (the average voltage of a pulse train), and an inertial section with a time constant τ (cf. Fig. 2a). We shall assume that the output meter M is inertialess and linearly reproduces the voltage obtained by subtracting the reference voltage U_0 that compensates the initial frequency component from the integrator voltage U_{out}. Assume that the average voltage U of the pulses is related to the frequency f by the coefficient c:

$$U = cf.$$

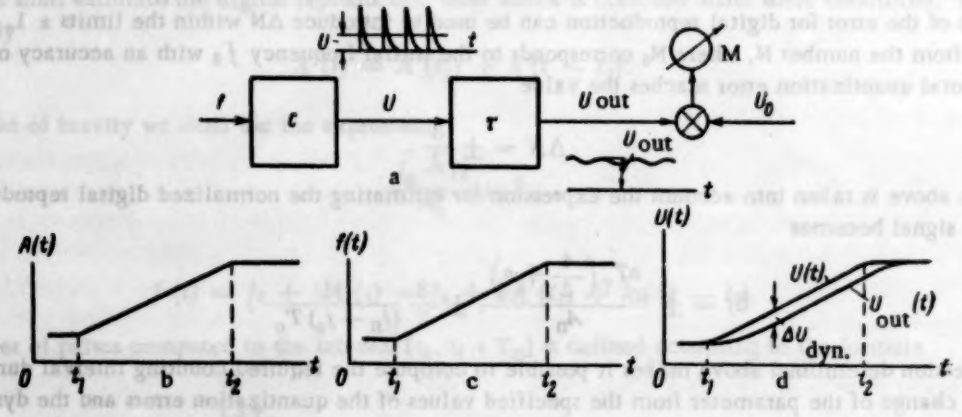


Fig. 2.

Then the relationship between U and the output parameter is expressed by the formula

$$U = cf_0 + ckA(t) = U_0 + ckA(t).$$

The choice of the time constant τ is governed by the necessity of obtaining sufficiently small pulsations of the output voltage U_{out}. It can be demonstrated that the amplitude of the pulsations is approximately equal to

$$U_{puls.} = \pm \frac{1}{2} \frac{c}{\tau}.$$

We shall normalize the pulsations to the working range of the meter scale by referring the voltage $U_{puls.}$ to the quantity $c(f_n - f_0)$. Then we shall obtain the normalized meter error caused by the pulsations

$$\delta_{puls.} = \pm \frac{1}{2(f_n - f_0)\tau}.$$

This expression is analogous to the expression for determining the quantization error in digital reception.

The inertial section reduces the pulsation that causes the appearance of the dynamic error associated with the lag that exists between the voltage U_{out} and the variations in the voltage U. For a linear rise of A (t) with a slope a (Fig. 2b) the steady-state value of the difference between U_{out} (t) and U (t) reaches the value

$$\Delta U_{dyn.} = -cka\tau.$$

This is illustrated in Fig. 2c. We shall find the normalized value of the dynamic error while simultaneously taking into account the change in its sign when U decreases with the slope a:

$$\delta_{dyn.} = \pm \frac{a\tau}{A_n}.$$

As we might expect, there is an analogy between the expressions for determining the dynamic errors for digital and analog reproduction. Assuming that

$$|\delta_{puls.}| = |\delta_{dyn.}| = \delta,$$

we find that

$$\tau = \frac{1}{2(f_n - f_0)\delta}, \qquad a = 2\delta^2 A_n (f_n - f_0).$$

Now it is possible to compare the two methods for receiving signals in a frequency or pulse-frequency telemetering system. We shall perform this for the "Chl" system of the Central Scientific Research Institute of Automation which has $f_0 = 4$ pulses/sec, $f_n = 20$ pulses/sec. We assume $\delta = 0.01$. In order to be specific we assume that we measure pressure with an upper limit of $A_n = 10$ kg/cm^2. Then for the case of digital reception with one channel (n = 1) we obtain $T_c = 9.4$ sec., $a = 0.007$ kg/cm^2 · sec. For the case of analog reception we have $\tau = 3$ sec., $a = 0.032$ kg/cm^2 · sec.

Thus, the choice of parameters for digital converters that are used for signals in analog telemetering systems must be made while taking into account the quantization errors and the dynamic error in accordance with the relationships derived in this paper. A comparison of the two methods for reproducing signals in a frequency system (the digital and analog methods) can be performed using the method presented here.

THE INVESTIGATION OF BRANCHED POWER NETS AS COMMUNICATION AND REMOTE CONTROL CHANNELS

A. A. Il'in (Kiev)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8, pp. 1088-1094, August, 1961

The author cites the results of investigating the electrical parameters of certain types of overhead and cable power distribution nets in the range 10 to 150 kc. An analysis is given of the propagation of electromagnetic energy over such lines, and methods are cited for computing them. Along with analytical computation methods the author presents recommendations on simulating high-frequency remote control communication channels. Certain information is given on the electrical noise in distribution power nets, and brief requirements are cited for the design of the remote control equipment.

The necessity of a broader introduction of existing means of remote control and the design of new units urgently poses the problem of communication channels.

In certain cases the use of existing branched power nets for large industrial enterprises as communication channels is adequate (electrical railway lines and subway lines, oil industry power lines, mine power systems, etc.). They are especially expedient to use underground where the application of radio communication is difficult.

1. The characteristics for Various Types of Power Nets

The investigation of certain types of branched power nets with voltage ratings of from 120 v to 6kv (both dc and ac) have demonstrated that these lines have a number of special features which distinguish them from conventional communication lines. Such features must include the specific nature of the kilometer parameters of the line which are caused by its construction, the presence of a large number of inhomogeneities along the lines which are caused by various power loads, transformers, etc., the presence of a characteristic mode of operation for an open line (the no-load mode), and the great variety of branched power net circuits with many branch lines whose length may be commensurate with the wavelength or a portion of a wavelength (Fig. 1).

In order to analyze the electromagnetic processes in branched power lines it is necessary to establish the value of the electrical parameters of uniform lines of varying design which make up these nets.

The table cites values of the wave impedance and the kilometer attenuation for certain types of power circuits in the frequency range 10 to 150 kc.

A mine contact net was analyzed in the greatest detail, both experimentally and theoretically.

It is evident from the table that the kilometer attenuation for the majority of lines is small. Therefore, when there are no additional losses along the line a transmission range of several tens of kilometers is guaranteed. The losses are appreciable only in flexible rubber cables of the "CRShS" type.

In branched power nets the overall working attenuation is determined not only by the kilometer attenuation but also by the attenuation b_i introduced to the line by various inhomogeneities. The attenuation b_i in certain cases may reach an appreciable value. A low-ohm load introduces especially large attenuation into the line.

The magnitude of the introduced attenuation is determined as a function of the ratio between the input load impedance and the wave impedance of the line (Z_l/Z_w).

It is evident from Fig. 2 that the value of b_i decreases as the ratio Z_l/Z_w increases. In practice b_i must be lowered to a value of the order of 0.1 to 0.2 nepers; this is assured (based on the example of high frequency power line channels) by introducing high-frequency suppressors which raise Z_i to the appropirate value.

Type of branched power net	Wave impedance, ohms	Kilometer attenuation, nepers/km	Circuit for the line	Remarks
A contact net for a mine	200 to 260	0.1 to 0.4	Single-wire	Connected equipment: trolley rails
	100 to 130	0.2 to 0.5	Two-wire	
A contact net for a railroad or a trolley line	300 to 400 150 to 200	0.05 to 0.2 0.1 to 0.3	Single-wire Two-wire	The same
A flexible cable of the type used in lines (type "CRShS")	70 to 80	0.2 to 2.0	Single-phase ground, three-phase ground	–
	50 to 60	0.3 to 2.5		
An armored cable of the type "SB" (copper strands)	30	0.1 to 0.2	Single-phase armored type, three-phase armored type	–
	20	0.1 to 0.2		

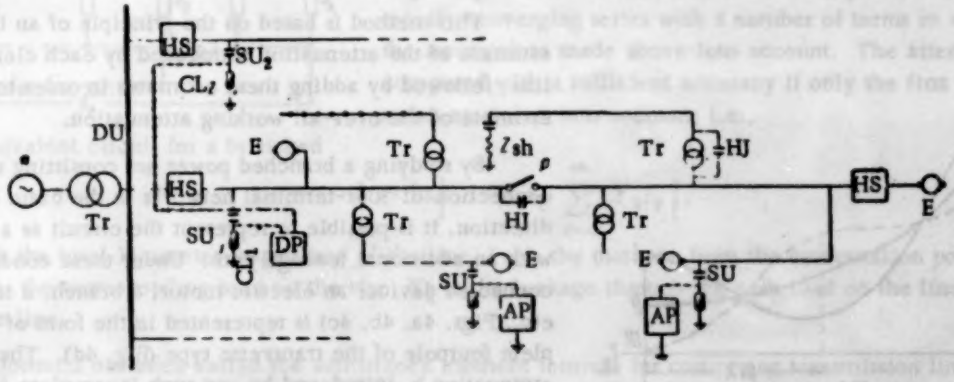


Fig. 1. Over-all diagram of the branched power net. DP, AP represent the dispatcher and actuating points; SU are the switching units; CL are the connecting lines; HS are the high-frequency suppressors; HJ are high-frequency jumpers; E are the electric motors; Tr are the transformers.

2. The Propagation of Electromagnetic Waves over a Branched Power Net

In investigating the nature of the propagation of electromagnetic energy over a branched power net it was established that the variation in line voltage is wavy in nature (Fig. 3). The basic cause for this is the formation of standing waves as a result of no-load operation of the line.

The intensity with which standing waves are formed in the branched power net are characterized by the standing wave coefficient k_{sw} whose magnitude depends on the value of the electrical parameters of the line.

The appearance of standing waves leads to instability in the operation of high-frequency channels, to an increase in the attenuation in the line, and to a fluctuation of its input impedance over wide limits. In order to eliminate standing waves, the ends of the line and the ends of long branches ($l_{br} > 1/8\ \lambda$) must be loaded by impedances (shunt) $Z_{sh} \approx Z_w$. It is most expedient to make the shunt in the form of a series RC circuit where C_{sh} is a blocking capacitor and R_{sh} assures the required loading of the line ($R_{sh} \approx |Z_w|$). The insertion of Z_{sh} at the line end adjusts it to an optimum traveling-wave mode; under these conditions the working attenuation is reduced and the input impedance is equalized.

3. Computing the Attenuation

The great variety of circuits and designs for branched power lines and the special features of these circuits considerably complicate the analysis and computation of high-frequency remote control channels.

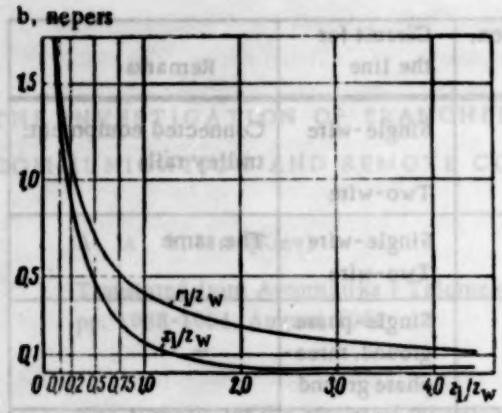


Fig. 2. Graph for the variation of the attenuation introduced for a purely resistive or purely reactive line inhomogeneity.

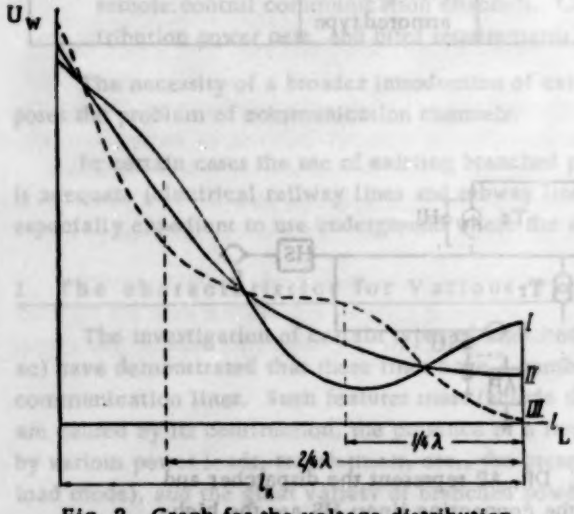


Fig. 3. Graph for the voltage distribution over the line. I) For an open end (no-load); II) for a matched load or very large losses in the line; III) for a short-circuited end (the short-circuit mode).

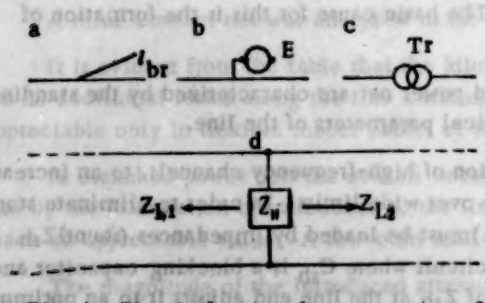


Fig. 4. Circuit showing the connection of a line inhomogeneity.

The use of existing computation methods in the theory of wire communication for these purposes is extremely difficult in view of the cumbersome and complex nature of the computations.

Therefore in determining methods for computing branched power lines it is natural to resort to approximate but simpler methods. In a number of cases it is quite adequate to compute merely the order of magnitude of the line attenuation or of the line input impedance.

In order to compute different types of lines we used two analytical methods of computation: the method of estimating the attenuation elements, and the method of admittance moments.

The use of one or the other method for performing the computation is determined as a function of the complexity of the line circuit and its length.

4. Methods for Estimating the Attenuation Elements of the Line

This method is based on the principle of an individual estimate of the attenuation introduced by each element of the line, followed by adding these estimates in order to obtain an estimate of the over-all working attenuation.

By studying a branched power net consisting of a complex connection of four-terminal networks in the basic (computed) direction, it is possible to represent the circuit as a two-wire line with an additional leakage path. Under these conditions any connected device: an electric motor, a branch, a transformer, etc., (Figs. 4a, 4b, 4c) is represented in the form of an incomplete fourpole of the transverse type (Fig. 4d). The working attenuation b_i introduced by any such incomplete fourpole consists of the input attenuation of the fourpole and the transmission attenuation. Its magnitude is determined by the inhomogeneity input impedances Z_i and the line impedance in either direction Z_{L1} or Z_{L2}. In order to simplify the computation formulas it is assumed that the outputs of the transmitters in the equipment are matched to the line; i.e., $Z_{L1} \approx Z_c$. Therefore,

$$b_i = \ln \left| \frac{Z_w + Z_{L2}}{\sqrt{Z_w Z_{L2}}} + \frac{\sqrt{Z_w Z_{L2}}}{2Z_i} \right| . \qquad (1)$$

Substituting the values of the impedances Z_{L2} and Z_i, it is possible to obtain the attenuation b_i introduced by any line connection.

In summary, the over-all working attenuation of a high-frequency channel for a branched power circuit is determined from the sum of all the attenuation components:

$$b = \beta_{km} L + \beta_{cl} l_{cl} + \sum_{q=1}^{m} b_{qi} + b_{add} . \qquad (2)$$

β_{km} and β_{cl} are the kilometer attenuation of the power circuit and the connecting line; L, l_{cl} are the lengths of the computed section of the circuit and the connecting line; m is the number of inhomogeneities in the line; b_i is the working attenuation introduced into the line by its various inhomogeneities; b_{add} is the additional attenuation.

Such a method for computing b is similar to the method for computing high-frequency channels which are based on electric power transmission lines and is simple and convenient for practical computations. The accuracy of the results depends on the circuit and the dimensions of the line. Good agreement between the computed and measured data is obtained for large lengths and a low degree of branching for the line.

In computing branched and short power nets this method does not give the required accuracy and becomes more cumbersome and difficult since in this case there is a strong effect produced by one line on the other because of the large number of inhomogeneities and the small distance between them; under these conditions the input impedance $Z_{L\Sigma}$ at any point is different and varies over wide limits. Therefore, in such lines the leakage introduced by various inhomogeneities must be treated while taking into account the nature of the leakage distribution along the line.

5. The Method of Line Admittance Moments

Approximate computation methods which have been widely used for computing city and rural wire radiobroadcast lines are of great interest. A considerable simplification of the computations can be achieved in the case under study by replacing the hyperbolic functions by the first one or two terms in their series; this is permissible for short transmission line sections.

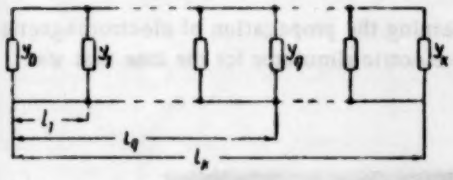


Fig. 5. Equivalent circuit for a branched power net.

Representing the circuit under study in the form of a ladder connection of fourpoles (Fig. 5), the ratio between the voltages at the beginning and end of such a line can be expressed in the form of a rapidly converging series with a number of terms m + 1 when we take the assumptions made above into account. The attenuation can be computed with a sufficient accuracy if only the first two terms of the series are taken into account; i.e.,

$$b = \ln \left| 1 + Z_L \sum_{q=0}^{m} Y_q l_q \right|, \tag{3}$$

where Z_L is the total kilometer impedance of the line, l_q is the distance from the computation point in the branched power net to the corresponding load on the line, Y_q is the leakage through the q-th load on the line, m is the number of loads on the line.

This formula has been called the admittance moment formula for computing transmission lines for wire radiobroadcasting; this formula is analogous to the expression $\sum_{q=0}^{m} Y_q l_q$ with a static first order moment. Basing ourselves on this, it is possible to introduce the concept of the coefficient ν for the distribution of the load admittance for the line. The coefficient ν, depending on the nature of the load distribution along the line, indicates the spot at which the basic load concentration occurs. A correct consideration of the quantity ν lowers the computation error to a considerable extent.

The expression for ν can be found by determining the admittance moment for the specified line which has a length l in terms of the admittance moment for an equivalent line which is terminated by a concentrated admittance:

$$\sum_{q=0}^{m} Y_q l_q = \nu l Y = l_1 Y. \tag{4}$$

From this it follows that

$$\nu = \frac{l_1}{l}. \tag{5}$$

The value of ν decreases when the load is concentrated toward the beginning of the line and increases when the load is shifted toward the end of the line; ν fluctuates within the limits 0 to 1.

In formula (4) the quantity νY is the total admittance of the line when the admittance distribution is taken into account. For convenience in computation we originate from the nature of the line admittances and subdivide the quantity νY into three groups: the resistive admittance Y_r, the reactive admittance Y_x, and the admittance Y_L of the line proper.

As a result formula (3) becomes

$$b = \ln|1 + Z_L L(\nu_r Y_r + \nu_x Y_x + \nu Y_L)|. \quad (6)$$

In computing the contact net for a mine the resistive admittance includes the electrical railway engines, the electrical transport sub-stations and the shunt admittances $Y_r = n/Z_{1L}$. The reactive admittance includes all short unprocessed branches whose input impedance is purely capactive $Y_x = \omega C \sum_{q=1}^{m} l_{br}$. The line admittance is determined by the kilometer parameters and for $G \ll \omega C$, $Y_L = \omega CL$. Therefore the computation formula for the attenuation is written in the following form in this case:

$$b = \ln\left|1 + Z_L L\left[\nu_r \frac{n}{Z_{1L}} + \omega C\left(\nu_x \sum_{q=1}^{m} l_r + 0{,}5L\right)\right]\right|, \quad (7)$$

where n, m respectively denote the number of resistive loads (equivalent to the input resistance of an electric locomotive) and reactive loads (branches); l_{br} is the length of a branch; ν_r ν_x are the distribution coefficients for the resistive and reactive loads; C is the kilometer capacitance of the line.

In order to determine the accuracy of the computation method we performed computations of several contact net circuits for mines; the results were in good agreement with experimental data.

6. Electrical Simulation of the Line

In order to verify the basic theoretical results and conclusions concerning the propagation of electromagnetic waves over complex branched power nets we designed and constructed an electric simulator for the line that was equivalent to a mining contact net (Fig. 6).

In this case electrical simulation provides extensive possibilities for studying various modes of operation of the line under conditions that are convenient for performing investigations.

The results of measurements performed using the simulator fully verified the basic physical picture of the phenomena which occur in such circuits: the formation of standing waves, the magnitude of the attenuation, etc. Therefore, in accordance with the recommendations of the author simulation has found practical application in designing high-frequency communication units based on contact nets in mines by "Ukrenergochermet".

Using simulation it is possible to determine the dependence of the line attenuation and the input impedance on the variation of the circuit parameters, to establish the optimal modes of operation for the line, to determine the spots at which to install the high-frequency channel equipment, etc.

7. Electrical Noise in Branched Power Nets

The intensity of electrical noise in branched circuits and the range of frequencies occupied by these signals are very great. We observe an especially high noise level in contact nets for electrical transportation. For example, in measuring the noise in a mine contact net by means of an "IP-13" noise meter the noise voltage in the range 20 to 150 kc was equal to 5 ± 40 mv or more. As the frequency increases the noise intensity drops.

The rectifiers, which produce harmonic components, represent one of the basic sources of noise in dc power circuits. Therefore, in order to suppress electrical noise all rectifier units must be specially constructed. For the same purpose it is necessary to construct the trolley suspensions and the track, etc. in such a way as to reduce sparking; this reduces the formation of pulse noise.

One of the basic sources of noise in a power cable net with a voltage rating of 127/380 v is welding aggregates which produce an especially high noise level in the investigated frequency range. Therefore all welding aggregates must be appropriately "treated". In general the noise voltage in these circuits is appreciably lower than in contact nets and amounts to 0.5 to 2 mv in the 5 kc band. Moreover, it is necessary to underscore the fact that the problem of investigating electrical noise in branched dc and ac power circuits requires further detailed study.

AUTOMATIC REMOTE CONTROL OF WATER DISTRIBUTION

I. B. Chizh (Tashkent)

(Presented before the Conference on Remote Control Science and Engineering in Moscow, November 21, 1959)
Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8, pp. 1095-1099, August, 1961
Original article submitted May 7, 1960

Problems of automatic remote control of irrigation canals are discussed. It is proposed to develop a special mathematical procedure for solution of the involved control problem on the basis of the theory of games. For solution of the differential equations describing the transient processes in a canal by means of electronic computers, it is recommended to use the method of characteristics. A solution is given for the problem of localization of a transient problem in a section of a canal.

The operation of an installation on one of the canals of an irrigation system can not be considered by itself, in isolation from the rest of the system, because a change in the operating conditions of one of the canals has an effect on the conditions in the other canals. This interaction is brought about by upstream propagation through the very canal where the action started. When remote mechanization of installations is introduced, there is a new bond imposed upon the irrigation system, which, in distinction from the one prevailing before, is of an artificial nature: the use of automation introduces a link between the conditions of the water movement in the channel and the operational characteristics of the installation. Obviously, the synthesis and analysis of a remotely controlled automatic irrigation system is impossible without exhaustive data on the static and dynamic properties of all elements included in the system and without knowledge of the mathematical relationships describing their characteristics.

At the present time it is considered advantageous to refrain from direct remote control of hydraulic installations on canals, and to carry out the control of their operations by changes in the settings of automatic regulators. However, further investigation leads to the conclusion that equipment of the water outlets with automatic regulators is insufficient and does not overcome all difficulties in the control of large irrigation canals. In some cases, when constant delivery automatic regulators are used, violations of normal operating conditions of a canal may take place, such as spilling over of water over a dam or inadequate supply to users at the end of the line. This is a consequence of the fact that the water outlets equipped with regulators are unable to increase the outflow when the canal is overfilled with water, and do not cause an increased intake of water when the water in the canal is low. To manage the canal operations correctly, it is necessary to be able to foresee the processes which will take place in the canal when the normal water flow in it has been disturbed.

The transient processes in a canal are described by a hyperbolic system of differential equations with partial derivatives as follows:

$$i_0 - \frac{\partial H}{\partial s} = \frac{U^2}{C^2 R} + \frac{\partial U^2}{2g\partial s} + \frac{\partial U}{g\partial t},$$

$$\frac{\partial F}{\partial t} + \frac{\partial Q}{\partial s} = 0, \qquad (1)$$

where i_0 is the longitudinal incline of the canal bottom, F the effective cross section area, H the stream depth, R the hydraulic radius, U the mean velocity of the liquid flow, s the distance on the longitudinal axis of a given section from the cross section used as the origin, g the gravitational acceleration, Q the quantity of water flowing per unit time through a given effective cross section, t the time, and C the Chezy coefficient.

It is well known that there is no general solution for the system of equations (1). However, it is entirely possible to use a fast computing machine for the solution of certain specific cases. One of such problems which it is feasible to solve by means of a computer is presented at the end of this paper.

Transfer of irrigation systems to dispatcher management by means of remote control was a big step forward. However, even now it is already possible to go further than introduction of automation and "distant mechanization". In a "distant mechanization" system, a human dispatcher is connected by means which are part of the distant control system with mechanisms and instruments located on the canal. At any moment the dispatcher can obtain information about the status of the water movements, and is able to interfere in the course of the processes involved. In times of rearrangement of the discharge regime, and also during emergencies, the dispatcher has to make very quick and responsible decisions. In such cases, "remote mechanization" proves to be of inadequate help. Thus, we are faced with the need for automatic distant control of irrigation canals. In this, the functions of the dispatcher could very well be carried out by a computing machine which as stated above has been recommended so far only for representation of transit processes in a canal.

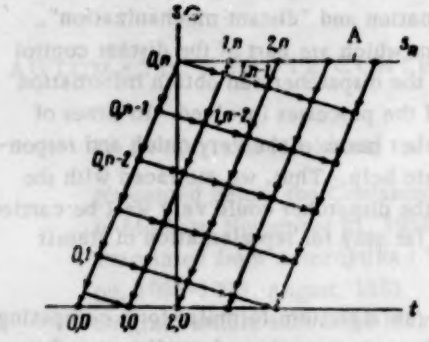
The mathematical apparatus for development of an operational procedure (an algorithm formula) for a computing machine acting as a dispatcher can be provided by the theory of games. Indeed, in the complex of conditions under which a dispatcher has to control an irrigation system, it is possible to see the same situation as in a game: The dispatcher aims to hold the supply of water to the strictest possible compliance with that planned. The "opponent" – objectively existing unfavorable circumstances – have the opposite aim. The dispatcher makes "moves" to minimize certain functions such as, for instance, the differences between the actual and planned water consumption figures; the opponent tries not to let that happen. The dispatcher has at his disposal several "strategies." The "opponent" also has available great possibilities: rise of air temperature, strong winds, unforeseen actions of the personnel servicing the system, accidents, etc. A success in the actions of the dispatcher can be regarded as his "gain" and as a "loss" of the opponent, and vice versa; and the gains and corresponding losses can be measured by the same figure with opposite signs (one with a plus, the other with a minus).

After the problem has been brought into the above form, it can be reduced to the most studied case of the theory of games – a two-person zero-sum game. The actions of the dispatcher represent a choice of one of the available strategies or, in other words, his personal, deliberate "moves." The opponent, in addition to personal, deliberate moves, makes also accidental or random moves; the time and probability of occurrence of a random move must be specified when the problem is set up.

The task of a machine replacing the dispatcher would consist in evolving the best strategy. The theory of games makes it possible to review the opportunities offered by a mechanical dispatcher. If the dispatcher before every move knows all preceding moves of his own and also of his opponent (which is a so-called game with full information), then there exists an optimal strategy; by using this strategy the dispatcher will be sure not to get any result which is worse than a certain predetermined result. There is, however, also another case possible: the distant control arrangement and the irrigation system are so designed that the dispatcher when making his more does not have full information about the preceding actions. For instance, there are two dispatcher points; both dispatchers are like two partners acting against a common opponent, but one of these dispatchers may not know what the other dispatcher has done before or is doing at the particular moment. In this case, an optimal strategy may not exist, that is, the outcome of the game is indefinite. The problem that the dispatcher has to solve can be modified in such a way that even in the case of indefinite games it will be possible to obtain feasible recommendations for an advantageous course of action. For this purpose, the problem should be set up in the following way: The dispatcher is not required to find the best strategy, but only to determine what the probabilities are that the available strategies will produce the most advantageous result. With this generalization of the problem, in the general case the probabilities of appearance of one or the other of the opponents strategies are also considered to be known. Apparently, there is good reason to believe that the theory of games can be used for the development of an algorithm (mathematical procedure) for computation of the operation of an automatic dispatcher in the control of an irrigation system.

In conclusion we will examine an example which confirms certain aspects put forward in this paper and before [2].

Let us assume that we have an irrigation canal, all water outlets from which are equipped with automatic regulators which maintain constant water supplies to the offtake canals. Let there be two kinds of regulators: the regulator of the first kind are those which only maintain constant discharges into the offtake canals and do nothing else; such regulators do not cause any replenishment of flow or head in the parent or main canal. To the second kind we relate such automatic regulators which maintain a certain constant head in the main canal, in consequence of which the discharges into the branch canals through standard orifices remain constant, considering that there is free outflow to the branch canals (negligible flow resistance).



Let us further assume that in the water outlet which is in the s_n − th sluice (the word sluice being used here in its most general sense − that of a device for checking and adjusting flow and head at a diversion point) at some moment of time a rearrangement of the discharged quantity of water must take place which proceeds according to the law $q_n = q_n(t)$. Prior to the end of the transient process in the canal caused by the mentioned discharge rearrangement, produced, e.g., by a resetting of the regulator in the s_n − th sluice, rearrangements also take place in the other water outlets located upstream from s_n, at certain moments of time and proceeding according to certain laws.

Let us undertake the task of determining the function $Q_h = Q_h(t)$ which must govern the change in water supply or consumption at the head of the canal in order that the regime of water flow in the main canal beyond s_n shall remain unaltered, that is, in a steady state. We shall do this considering that the following are given: the characteristic parameters of the water motion in the established state of flow which preceded the start of the discharge rearrangements, the laws according to which the water discharges in all the outlets are being changed, and all essential parameters of the canals themselves such as the incline of the bottom i_0, the profile of the canal, and the Chezy coefficient C.

As was stated above, the transient processes in canals are described by the system of equations (1). We shall carry out a numerical solution of this system by the method of characteristics.

Let us divide the section of the canal from the head to the s_n − th sluice into n equal parts, and let us determine the points $(t_{0,k};\ s_{0,k})$ which are situated on the initial characteristic (see figure). The coordinates $t_{0,k}$ are determined by means of the recursion formula:

$$t_{0,k} = t_{0,k+1} - \frac{2s_n}{\sum_{j=k}^{n} (U_{0,j} + \omega_{0,j})} \quad (k = n-1,\ n-2,\ \ldots,\ 1,\ 0). \tag{2}$$

The coordinate $s_{0,k}$ is determined as follows:

$$s_{0,k} = \frac{ks_n}{n} \quad (k = n,\ n-1,\ \ldots,\ 1,\ 0). \tag{3}$$

Here and further on the subscripts indicate to which point of the network the given parameters belong.

In expression (2)

$$t_{0,n} = 0, \qquad \omega_{0,k} = \sqrt{\frac{gF_{0,k}}{B_{0,k}}},$$

and B is the width of the "live" cross section, that is, the one actually filled with water.

The state which prevailed prior to the rearrangement of the discharges is determined by data contained in the setup of the problem. Thus, the values of the depth H, and of the mean velocity U from the initial characteristic are known; and the following quantities needed for further computations can be determined from known data:

$$\omega = \sqrt{\frac{gF}{B}}, \quad \lambda = \int \sqrt{\frac{g}{BF}}\, dF.$$

The values of ω, λ, and U on the halflength straight line $s = s_n$ can be considered to be known quantities, since they can be determined from the water consumption Q_n and the depth of the water H_n in canal n. Knowing the values of these parameters on the halflength straight line $s = s_n$ and on the line of initial characteristics (abscissa axis), we can construct a network representing the intermediate characteristics. We draw straight lines, as shown in the figure, through the points 1, n and 0, n which represent respectively the direct and the reciprocal (or inversely proportional) characteristics, and determine to a first approximation the crossing point coordinates $t_{1,n-1}$ and $s_{1,n-1}$ of these lines and the values of U and λ at these points, which are as follows:

$$t_{1,n-1} = t_{1,n} + \frac{s_{0,n} - s_{1,n} + (t_{1,n} - t_{0,n})\,\Omega_{av}}{W_{av} - \Omega_{av}}, \tag{4}$$

$$s_{1,n-1} = s_{1,n} + W \quad (t_{1,n-1} - t_{1,n}), \tag{5}$$

$$\lambda_{1,n-1} = \lambda_{1,n} - \frac{U_{0,n} - U_{1,n} + \delta_{2av} - \delta_{1av} + \lambda_{1,n} - \lambda_{0,n}}{2} \tag{6}$$

$$U_{1,n-1} = U_{1,n} - (\lambda_{1,n-1} - \lambda_{1,n}) + \delta_{1av} \tag{7}$$

where

$$W_{av} = U_{1,n} + \omega_{1,n}, \qquad \Omega_{av} = U_{0,n} - \omega_{0,n},$$

$$\delta_{1av} = g\left(i_0 - \frac{U^2}{C^2R}\right)_{1,n} (t_{1,n-1} - t_{1,n}),$$

$$\delta_{2av} = g\left(i_0 - \frac{U^2}{C^2R}\right)_{0,n} (t_{1,n-1} - t_{0,n}).$$

The second approximations are found by means of the same formulas as above. Only for W_{av}, Ω_{av}, δ_{1av}, and δ_{2av}, now the following more exact values are substituted:

$$W_{av} = \frac{1}{2} \sum_{j=n-1}^{j=n} (U_{1,j} + \omega_{1,j}),$$

$$\Omega_{av} = \frac{1}{2} (U_{0,n} - \omega_{0,n} + U_{1,n-1} - \omega_{1,n-1}),$$

$$\delta_{1av} = g\left(i_0 - \frac{Q_{1,n}^2 + Q_{1,n}Q_{1,n-1} + Q_{1,n-1}^2}{3K_{1\,av}^2}\right)(t_{1,n-1} - t_{1,n}),$$

$$\delta_{2av} = g\left(i_0 - \frac{Q_{0,n}^2 + Q_{0,n}Q_{1,n-1} + Q_{1,n-1}^2}{3K_{2\,av}^2}\right)(t_{1,n-1} - t_{0,n}),$$

where K_{1av} and K_{2av} are the average values of the discharge quantity modulus respectively in the section of the direct characteristic between the points (1, n) and (1, n − 1), and in the section of the reciprocal characteristic between the points (0, n) and (1, n − 1).

The coordinates of the other crossing points of the network lines and the values of U and λ for them are found in a similar manner; there will be some minor differences only in the determination of U and λ at the sluices in which water discharges are present.

Thus we have solved our problem which represents a mixed Cauchy-Rieman problem of equations of the hyperbolic type.

After the values of U and λ on the abscissa axis have been determined (which corresponds to the sluice in which the head junction is located) we can determine the function $Q_h = Q_h(t)$ which we were seeking.

The anticipation time, that is, the time by which the start of a rearrangement in the s_n−th sluice must be forestalled by a regulation at the head of the canal, is equal to the absolute value of $t_{0,0}$.

The relative simplicity of the mathematical procedure, and at the same time the cumbersomeness of the operations involved confirms the possibility and the advantage of using electric computers for the control of irrigation canals.

The example examined by us shows also that by means of computing machines it is possible not only to determine the results of certain actions of a dispatcher, but also to prescribe the operations which must be carried out by a dispatcher in order to secure a desired results.

LITERATURE CITED

1. M. D. Chertousov, Hydraulics [in Russian], State Power Press (1957).
2. I. B. Chizh, Automation and Mechanization of Irrigation Systems, and Transient Processes in Canals [in Russian] Transactions of the [SANIIRI] Central Asia Scientific Research Institute of Irrigation No. 95 (1958).

ANALYTIC AND STRUCTURAL DESCRIPTION OF MECHANICAL TRANSMISSIONS IN AUTOMATIC CONTROL SYSTEMS WITH RESTRICTIONS AND BACKLASH

Ts. G. Litovchenko and Yu. P. Yakovenko (Moscow)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 1100-1107, August, 1961
Original article submitted June 1, 1960

We consider a method which may be used to describe the backlashes and restrictions in mechanical transmissions used in automatic control systems. The method is based on an account of the elastic properties of transmission elements. The method is simple and can be applied both to grapho-analytical design and for simulation.

We propose ways of taking into account the energy dissipation resulting from the clash of mechanical transmission elements.

In the analysis of the operation of automatic control systems containing mechanical transmissions it is necessary in a series of cases to account for the presence of backlash and of various restrictions. Practice shows that backlashes are especially dangerous in loaded transmissions where the load inertia and elastic properties of the kinematic elements may prove to be the determining factors in the resulting self-oscillations.

There are at the present time two mathematical methods of describing backlash. The first method represents the backlash characteristics in the form of a hysteresis loop [1] for a totally unloaded transmission and can be called a kinematic means for describing backlash. It can only be used in the description of various measuring circuits in which the inertial force is negligibly small in comparison with the frictional forces.

The second method, which is described in [2, 3], is based on the description of the individual and simultaneous (joint) motion of the driving and driven elements by means of the equations

$$(J_1 + J_2)\ddot{\varphi}_1 = (J_1 + J_2)\ddot{\varphi}_2 = M_1 + M_2, \tag{1}$$

$$J_1\ddot{\varphi}_1 = M_2, \quad J_2\ddot{\varphi}_2 = M_2. \tag{2}$$

The transition from the equations of simultaneous joint motion (1) to the equations of individual motion (2), and vice versa, occurs at the time when

$$|\varphi_1 - \varphi_2| = \epsilon, \tag{3}$$

where ϵ is the half-interval of the backlash. Therefore, in order to assure uninterrupted motion we must impose correct initial conditions.

During the transition from simultaneous to individual motion the initial conditions are given by

$$\dot{\varphi}_{10} = \dot{\varphi}_{20} = \dot{\varphi}_1, \qquad \varphi_{10} = \varphi_{20} \pm \varepsilon = \varphi_1, \tag{4}$$

where φ_{10} and $\dot{\varphi}_{10}$ denote the initial coordinates and velocity of one of the bodies at the moment of separation. The transition from separate to simultaneous motion is accompanied by a blow (or shock); this significantly complicates the determination of the initial conditions for the simultaneous motion. In [3] the proposal is made that we assume that at the moment of contact of the two bodies $\dot{\varphi}_2$, the velocity of the driven body, be compared artificially with $\dot{\varphi}_1$, the velocity of the driver. Obviously this is true only when the acceleration of the driven body, due to the action of the load, is very large and its mass is negligibly small in comparison with that of the driver. In this case the driven body is almost instantaneously compressed under the action of the external force imposed by the driver and it starts to move together with the driver. This situation practically never exists in load transmissions.

When a proper appraisal is made of the dynamic phenomena accompanying the collision of two bodies the formulas used for the determination of the initial conditions of simultaneous (joint) motion become significantly more complex. In order to determine the initial conditions we must take into account the plastic properties of the bodies. We can obtain these properties if we examine the two limiting cases. In the first case the bodies seem to be stuck together after the collision and they move on further together. According to the law of conservation of momentum the mutual velocity $\dot{x}$ of the bodies after the collision is given by

$$\dot{x} = \frac{m_1\dot{x}_{10} + m_2\dot{x}_{20}}{m_1 + m_2}. \tag{5}$$

In the second case the bodies move away from each other after the collision. Since the kinetic energy and momentum are preserved we get

$$\frac{m_1\dot{x}_1^2}{2} + \frac{m_2\dot{x}_2^2}{2} = \frac{m_1\dot{x}_{10}^2}{2} + \frac{m_2\dot{x}_{20}^2}{2}, \quad m_1\dot{x}_1 + m_2\dot{x}_2 = m_1\dot{x}_{10} + m_2\dot{x}_{20}.$$

The velocity of both bodies after the collision is obtained from the formulas

$$\dot{x}_1 = \frac{m_1 - m_2}{m_1 + m_2}\dot{x}_{10} + \frac{2m_2}{m_1 + m_2}\dot{x}_{20}, \quad \dot{x}_2 = \frac{m_2 - m_1}{m_1 + m_2}\dot{x}_{20} + \frac{2m_1}{m_1 + m_2}\dot{x}_{10}. \tag{6}$$

Thus, when describing backlash by means of the equations for individual (2) and joint (1) motion we must include supplementary formulas that enable us to determine the initial conditions. This greatly complicates the modeling of dynamic systems with backlash. Besides, this type of approach does not permit us to make use of graphical-analytic methods based upon harmonic balance [1] and harmonic linearization [4].

All of these facts force us to seek a more easily applied method for the description of backlash and restrictions. In this paper an effort is made to give a simple structural description of backlash and restrictions taking into account all of the basic peculiarities involved in the collision of two bodies and the elastic properties of mechanical transmission having backlash and restirctions. The proposed method provides a simple way for performing graphical and analytic calculations and modeling.

1. Analytic and Structural Representation of Backlash and Restrictions

We can describe the behavior of a mechanical system with backlash and restrictions much more simply and yet with sufficient accuracy if we use the elastic properties of the system. Let us consider a system consisting of several masses separated by elastic elements. The forces or moments transmitted from one mass to the other are functions (nonlinear in the general case) of the deformation of the elastic elements separating these masses. These functions can usually be linearized if we put them in the form

$$M = k_y \delta, \tag{7}$$

where M is the moment transmitted by the system, δ is the magnitude of the elastic deformation, and k_y is the coefficient of elasticity; k_y depends upon the material from which the transmission elements are made, their dimensions, and form. The value of k_y can be determined experimentally or it can be calculated if we known the geometry of the transmitting elements.

Taking into account the elastic moment of rotational motion of two bodies separated by backlash, we can write the equations

$$J_1\ddot{\varphi}_1 = M_1 - M, \quad J_1\ddot{\varphi}_2 = M_2 + M, \tag{8}$$

where M is the moment of the reaction of one body upon the other as calculated by means of the formula (7); M_1 and M_2 are the total moments due to all the external forces (except the reaction moment) applied respectively to the first or second body. The deformation δ in the presence of backlash having a width 2ϵ is determined by the obvious relations

$$\delta = \begin{cases} \varphi_1 - \varphi_2 - \varepsilon & \text{where } \varphi_1 - \varphi_2 > \varepsilon, \\ 0 & \text{where } |\varphi_1 - \varphi_2| \leqslant \varepsilon, \\ \varphi_1 - \varphi_2 + \varepsilon & \text{where } \varphi_1 - \varphi_2 < -\varepsilon. \end{cases} \tag{9}$$

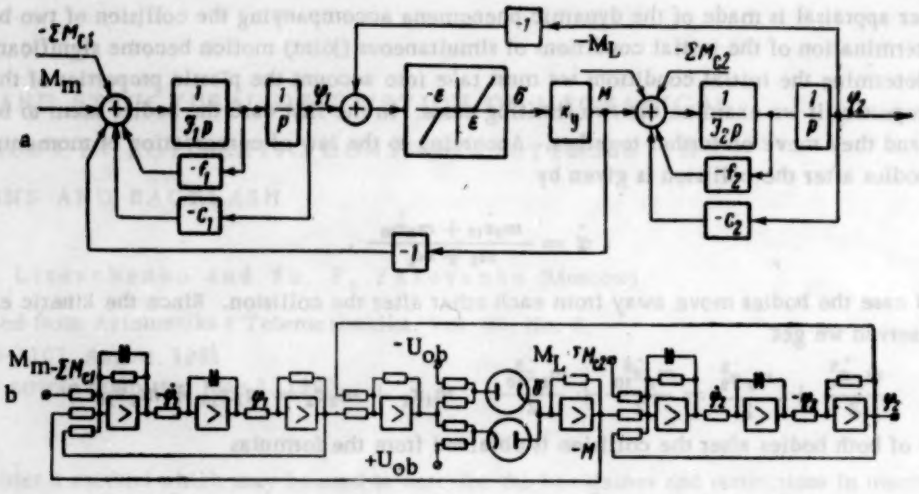


Fig. 1.

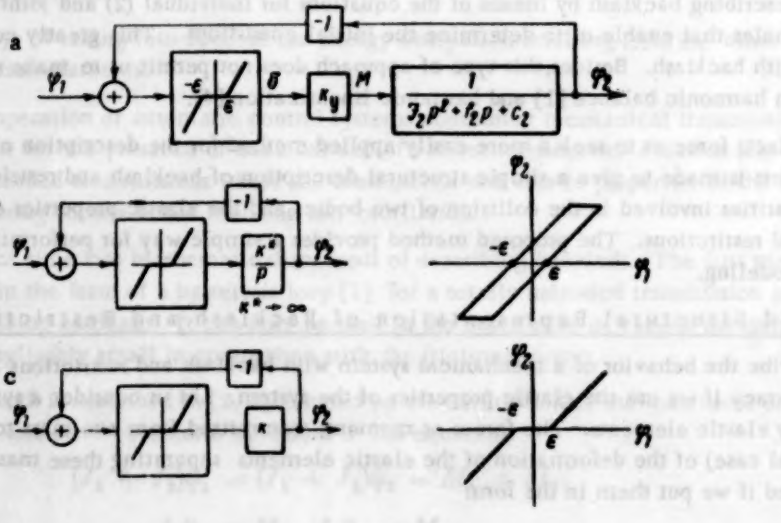


Fig. 2.

We can write equation (8) in a more detailed manner by separating the resistance moments which are proportional to the masses of the displaced bodies which are often encountered in the transmission of forces

$$J_1\ddot{\varphi}_1 + f_1\dot{\varphi}_1 + c_1\varphi_1 = M_m - M - \Sigma M_{c1}, \tag{10}$$

$$J_2\ddot{\varphi}_2 + f_2\dot{\varphi}_2 + c_2\varphi_2 = M - M_L - \Sigma M_{c2}. \tag{11}$$

In these equations f_1 and f_2 are the coefficients of the moments analogous to the case of viscous friction; c_1 and c_2 are the coefficients of the moments associated with the shift in positions; M_m is the moment associated with the motion, applied to the driver; M_L is the moment associated with the load that is applied to the driven element, and ΣM_{c1} and ΣM_{c2} are additional possible resistance moments.

A block diagram showing the structure of the system described by equations (7), (9), (10), and (11) is given in Fig. 1a. The figure is simple enough for graphical-analytic calculations and for modeling. The setup used for modeling the system is shown in Fig. 1b. In contrast to the set-up shown in [2] it does not require any changes or adjustments or calculation of the initial conditions in order to model the transition from individual to joint motion. When the dependence of the moment M upon the deformation δ is nonlinear the block diagram remains the same except for the fact that the characteristics of the nonlinear unit are curves rather than straight lines; this is required by the nonlinear dependence.

The use of a hysteresis loop to describe the backlash characteristic follows from Fig. 1, it is a particular case of this figure. The dependence of the position of the driven body φ_2 upon that of the driver φ_1 is described by means of equations (7), (9) and (11). The block diagram of Fig. 2a (for $\Sigma M_{c2} = 0$ and $M_L = 0$) is based upon these equations. If the moments of position and inertia are absent, then for a transmission with an infinitely large stiffness the block diagram describes the backlash in the form of a hysteresis loop (Fig. 2b). If the inertial and damping moments are negligibly small then the block diagram shows a zone of instability (Fig. 2c). The positional moment in this case, so to speak, "chooses" the backlash. The angle of inclination (the slope) of the branches which constitutes the characteristic of the instability zone is equal to the arc tg $k/1 + k$, where $k = k_y/c_2$.

The restriction on the motion (deviation) of the mechanical transmission elements in an automatic regulating or control system can be considered as a particular case of backlash. In this case the driven body is retarded ($\varphi_2 = 0$) and the magnitude of the gap ϵ depends upon the maximum value of the boundary magnitude φ_{1max}. The motion of a mechanical transmission with restrictions is described by equations (7), (9), and (10) if we substitute $\varphi_2 = 0$ and $\epsilon = \varphi_1 max$. The block diagram of an inertial element with restrictions is given in Fig. 3a. The nonlinear characteristic of the unstable zone is in the feedback loop. The block diagram can be constructed in such a manner that the nonlinear element is located in the direct path. It follows from equation (10) that

$$M_m - \Sigma M_{c1} - J_1\ddot{\varphi}_1 - f_1\dot{\varphi}_1 = M_0. \tag{10a}$$

where

$$M_0 = c_1\varphi_1 + M. \tag{12}$$

Using equations (7) and (9) we get

$$M = \begin{cases} k_y(\varphi_1 - \varphi_{max}) & \text{if} \quad \varphi_1 > \varphi_{max} \\ 0 & \text{if} \quad |\varphi_1| \leqslant \varphi_{max} \\ k_y(\varphi_1 + \varphi_{max}) & \text{if} \quad \varphi_1 < -\varphi_{max}. \end{cases} \tag{13}$$

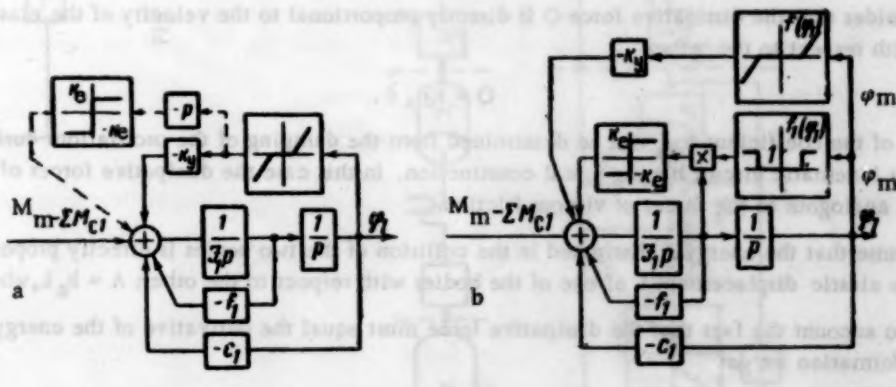


Fig. 3.

Equations (12) and (13) determine the connection between the change in φ_1 and the moment M_0 which is shown graphically in Fig. 4a. The slopes α_1 and α_2 which characterize this bond are functions of the coefficients c_1 and k_y

$$\alpha_1 = \text{arc tg}\,\frac{1}{c_1}\,; \quad \alpha_2 = \text{arc tg}\,\frac{1}{c_1 + k_y}\,.$$

In the particular case where restrictions are placed upon the motion of a system with infinitely large stiffness ($k_y \to \infty$) the graph of Fig. 4a degenerates into the customary nonlinear characteristic for this type of restriction as shown in Fig. 4b. In this form it may be used in graphical-analytic calculations used for automatic control systems.

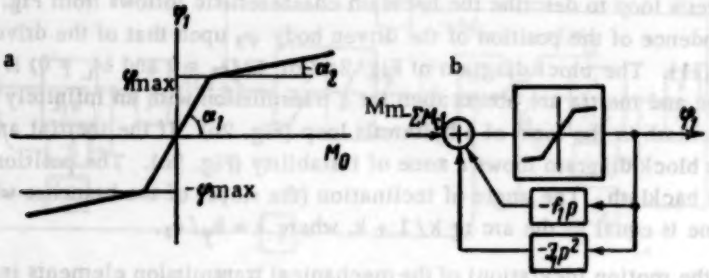


Fig. 4.

2. How to Take Account of the Inner Dissipative Forces

The methods described above for obtaining a structural representation of mechanical systems with restrictions and backlash do not take into account the energy loss which occurs in actual systems during collisions, i.e. the transformation of part of the kinetic energy into heat energy. This energy dissipation due to inner dissipative forces is associated with the elastic imperfections of the material of which the colliding bodies are composed and of the entire kinematic circuit, i.e., with the plastic deformation of the bodies at their point of contact and with the inner friction in the metal during elastic deformations of the system. If we take into account the energy dissipation we obtain a damping of the oscillatory motion of the bodies during collisions even when no external (with respect to the moving bodies) dissipative forces such as viscous or dry friction are present.

An exact picture of the energy dissipation during collision is extremely complex. However, we find that in order to obtain a structural (block diagram) representation which we can use as a model of an automatic control or regulating system where backlash is present in the mechanical transmission and where the mechanical system is subject to restrictions, it is possible to obtain accurate results with a simplified block diagram. We assume that there is no plastic deformation of the colliding bodies and that the kinematic circuit is subject only to elastic deformation. As far as the engineering approach is concerned the following two ways of taking account of the dependence of the dissipative forces upon elastic deformation seem to be the most logical.

1. We consider that the dissipative force Q is directly proportional to the velocity of the elastic displacement δ of one body with respect to the other

$$Q = k_{dis}\,\dot{\delta}.$$

The value of the coefficient k_{dis} can be determined from the damping of the oscillations during the collision of the elements in a kinematic circuit having a real construction. In this case the dissipative forces of internal friction in the material are analogous to the forces of viscous friction.

2. We assume that the energy A dissipated in the collision of the two bodies is directly proportional to the absolute path of the elastic displacement L of one of the bodies with respect to the other: $A = k_e L$, where $L = \int |\dot{\delta}|\,dt$.

Taking into account the fact that the dissipative force must equal the derivative of the energy A with respect to the elastic deformation we get

$$Q = \frac{d}{d\delta} A = k_e \frac{d}{d\delta} \int |\dot{\delta}|\,dt = k_e \frac{d}{d\delta} \int \dot{\delta}\,\mathrm{sign}\,\dot{\delta}\,dt,$$

or $Q = k_e \sin \dot{\delta}$.

In this case the inner dissipative forces are analogous to the forces of dry friction. When $\dot{\delta} = 0$ there is no dissipative force.

We may use a similar method to take into account the energy dissipation not only during backlash and restrictions but, in general, during elastic deformation of the elements in automatic control systems.

In Fig. 3a we show a block diagram of a mechanical transmission which is subject to restrictions; the dashed lines indicate the elements which take into account the dissipative forces in accordance with the second method outlined above.

In order to avoid the inclusion of a differentiating unit, this being undesirable for modeling, we will express the velocity of the elastic deformation by means of the relation

Fig. 5.

Fig. 6.

$$\dot{\delta} = \frac{d}{dt} f(x) = \frac{df(x)}{dx} \frac{dx}{dt} = \dot{x} f_1(x),$$

where $f_1(x)$ is a relay characteristic The block diagram corresponding to the expression is shown in Fig. 3b.

3. Examples

We would like to present some example demonstrating the use of block diagrams for automatic control or regulating systems containing backlash.

Example 1. Pneumatic servomechanism with a final control element that possesses inertia and where backlash is present in the transmission. Account is taken of the stiffness of the transmitting elements and of the enregy dissipation in the backlash (Fig. 5a).

The equation of motion has the form

$$F(t-\tau) = k_F [z(t) - k_{oc} x_1(t)],$$

$$m_1 \ddot{x}_1 + f_1 \dot{x}_1 = F - F_y,$$

$$F_y = k_y S, \quad F_\pi = F_0 \operatorname{sign} \dot{S},$$

$$S = \begin{cases} x_1 - x_2 - \varepsilon & \text{if} \quad x_1 - x_2 > \varepsilon, \\ 0 & \text{if} \quad |x_1 - x_2| \leqslant \varepsilon, \\ x_1 - x_2 + \varepsilon & \text{if} \quad x_1 - x_2 < -\varepsilon, \end{cases}$$

where z is the control signal, F is the force applied to the piston rod, k_F is the slope of the servomotor characteristic, k_{oc} is the slope of the feedback loop of the servomotor, τ is the delay in the pneumatic system, x_1 is the displacement of the motor piston rod, m_1 and m_2 are the masses of the piston and load respectively, f_1 and f_2 are the coefficients of viscous friction on the piston rod and load respectively, F_y is the force of elasticity, k_y is the coefficient of elasticity, S is the elastic deformation, F_L is the resisting force applied to the load, F_g is the dissipative force, with modulus F_0.

The corresponding block diagram is shown in Fig. 5b.

Example 2. An electric motor loaded by a system of masses which are connected by means of elastic elements and backlashes.

Let us examine the system which is shown schematically in Fig. 6a. Corresponding to this system we have the mechanical system: electric motor-reducer-final control element, where account is taken of the gaps and of the toothed transmission. The dashed lines in Fig. 6a separate the homogeneous blocks, of which there can be any number.

The equation of the motor is given by

$$\omega_m = F(U, M_m), \tag{14}$$

where U is the voltage applied to the motor and M_m is the moment at the output shaft of the motor.

The equations for one of the blocks are given below:

$$\omega_1' = i_1 \omega_m, \quad M_m = i_1 M_1', \quad M_1' = k_{y1} \delta_1,$$

$$\delta_1 = \begin{cases} \varphi_1' - \varphi_1 - \varepsilon_1 & \text{for} \quad \varphi_1' - \varphi_1 > \varepsilon_1, \\ 0 & \text{for} \quad |\varphi_1' - \varphi_1| < \varepsilon_1, \\ \varphi_1' - \varphi_1 + \varepsilon_1 & \text{for} \quad \varphi_1' - \varphi_1 < -\varepsilon_1, \end{cases} \tag{15}$$

$$\dot{\varphi}_1' = \omega_1', \quad \dot{\varphi}_1 = \omega_1, \quad J_1 \dot{\omega}_1 = M_1' - M_1 - M_{1c}.$$

Equations analogous to (15) may be written for the other blocks. The corresponding block diagram is given in Fig. 6b.

LITERATURE CITED

1. The Basis of Automatic Control Theory [in Russian]. Edited by V. V. Solodovnikova, Mashgiz (1954).
2. Onufriyuk, S. P. and Fel'dbaum, A. A., Electronic model of backlash, Avtomatika i Telemekhanika, 17, No. 6 (1956).

3. Kogan, B. Ya., On the modeling of automatic control systems having typical nonlinear characteristics, Avtomatika i Telemekhamka 16, No. 2 (1955).
4. Popov, E. P., Dynamics of Automatic Control Systems [in Russian] Gostekhizdat (1954).

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

ESTIMATING THE RELIABILITY OF AUTOMATIC SYSTEMS FROM THE RESULTS OF TESTING AN INCOMPLETE SET OF EQUIPMENT

S. M. Kuznetsov (Moscow)

Translated from Avtomatika i Telemekhanika Vol. 22, No. 8,
pp. 1108-1116, August, 1961
Original article submitted November 21, 1960

The paper studies a method for estimating the reliability of systems on the basis of results obtained by testing an incomplete set of equipment.

The dependence of the estimate error on the structure of the system is established along with the relationship between the volumes of the element groups which enter into the composition of the tested and untested units.

A statistical estimate of the reliability of automatic systems is associated with testing a large amount of equipment (the power of the vibration stands, the volume of the refrigeration chambers, the humidity, the heat, etc.).

Systems consisting of a large number of complex and cumbersome blocks and aggregates must in that case be tested separately; under these conditions a portion of the most cumbersome blocks and aggregates is not tested.

The absence of the necessary data on the reliability of the individual parts of the system substantially lowers the accuracy and reliability of the estimates and often eliminates the practical value or expediency of such test. Under these conditions it becomes necessary to develop more refined methods for estimating the reliability of systems from the results of testing an incomplete set of equipment.

General Concepts for Solving the Problem

A distinguishing feature of modern automatic systems is their use of a large number of elements, sections, and modules of the same type which operate under approximately identical and almost optimal operational conditions. These special features make it possible to use the differentiated method of statistical estimation of system reliability in solving the problem posed above while taking into account the relative number of failures in elemens and sections (modules) of the same type in the system.

The essence of the differentiated method of obtaining an estimate is based on the following assumptions.

If the system consists of k different groups of elements of the same type which are characterized by approximately the same damage probability, then the over-all damage probability P_s for the system (with respect to all groups of elements of the same type) can be expressed in terms of the damage probability P_{2i} for the system with respect to each group of elements of the same type:•

$$P_s = 1 - (1 - P_{21})(1 - P_{22}) \ldots (1 - P_{2k}) \qquad (1)$$

For $P_{2i} \ll 1$ eq. (1) can be written in the following simpler form with an error up to erms of second order smallness•• :

$$P_s \approx \sum^{k} P_{2i}. \qquad (2)$$

• It is assumed that damage to any element in the system is associated with damage to the system.

•• The methodological error involved in the estimate for this case can be determined analytically from the formula

$$\Delta P_s = -\Sigma P_{2i}P_{2j} + \Sigma P_{2i}P_{2l}P_{2l} - \ldots + (-1)^k P_{2i}P_{2j}P_{2l}\ldots P_{2k}. \qquad (A)$$

For an experimental estimate of the quantity P_{2i} we specify the statistical characteristics of the reliability indices for the systems: the mean-arithmetic frequency of system failures with respect to each group of elements a_{2i}, and the statistical dispersion D^2_{2i}. These statistical characteristics converge with respect to probability to the mathematical expectation a_{2i} and the mathematical dispersion D_{2i} of the reliability indices as the number of samplings increases*.

The numerical characteristics of the system reliability indices a_s and D_s can be expressed in terms of the quantity $a_{2i}D_{2i}$.

In accordance with (1) we shall have **

$$a_c = \sum^{k} a_{2i},$$
$$D_c = \sum^{k} D_{2i} \qquad (2a)$$
$$\sigma_c = \sqrt{\sum^{k} \sigma^2_{2i}}.$$

In analogous fashion we can establish the relationship between the damage probabilities P_{2i} for the system and the damage probability P for its elements.

Assuming that

$$P_{2i} \approx \sum^{M_i} P_e \qquad (3)$$

we have by analogy with (2a):

$$a_{2i} = \sum^{M_i} a_e, \; D_{2i} = \sum^{M_i} D_e, \; \sigma_{2i} = \sigma_e \sqrt{M_i}, \qquad (3a)$$

where M_i is the number of elements of the same type in the group.

Substituting the value of σ_i from (3a) into (2a), we obtain

$$\sigma_s = \sqrt{\sum^{k} M_i \sigma^2_e}. \qquad (3b)$$

In accordance with (3b) the mean square deviation of the reliability indices for the system from their average probability values depends not only on σ_e but also on the structure of the system: the number M of elements of the same type and the number of groups k.

Relationships (2) and (3) can be used for a statistical estimate of system reliability from the results of testing an incomplete set of equipment.

---

* System reliability, depending on the purpose of the system and the conditions governing its operation, is characterized by various indices: the reject fraction (the relative number of faults in the system during fixed time interval), the intensity or frequency of failures, etc. Irrespective of the established indices, the estimate of system reliability is usually associated with determining the absolute or relative number of faults for the system over fixed time intervals.

** Equations (2a) are based on the throrem for the mathematical expectation and dispersion of a sum of independent random quantities.

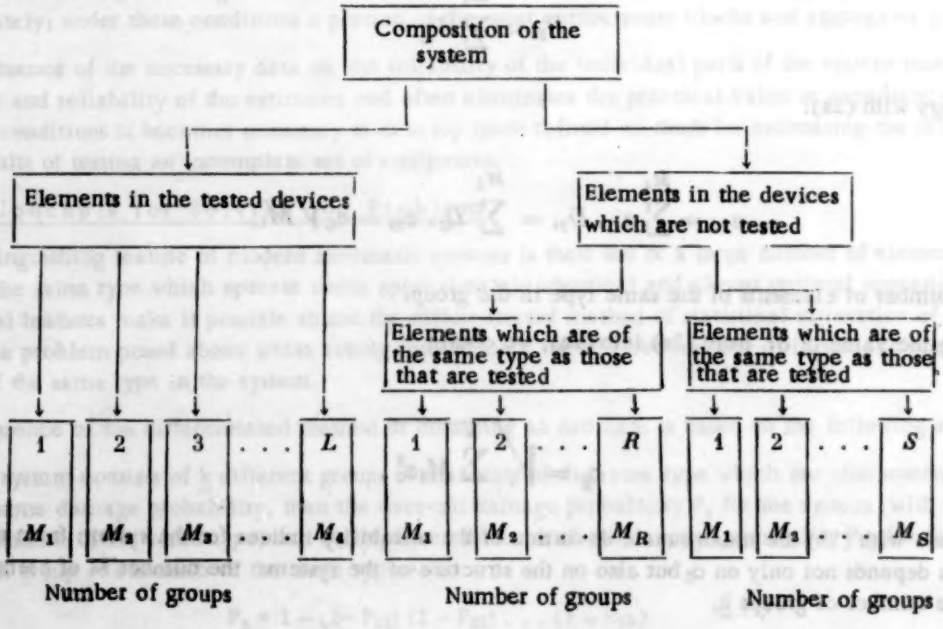
We shall study the effectiveness of using the differentiated method of estimation in solving the problem posed above.

If system elements of the same type are subdivided into groups while taking into account the possibility that they are tested within the context of the devices and blocks of the system (cf. figure), then the over-all probability P_s of system damage (with respect to all the complexes of groups) can in analogy with (2) be expressed approximately in terms of the probability of damage to the system with respect to each complex of groups P_{c1}, P_{c2}, and P_{c3}:

$$P_s \approx P_{c1} + P_{c2} + P_{c3} \approx \sum^{L} P_{ti} + \sum^{R} P_{nj} + \sum^{S} P_{sl}, \tag{4}$$

where P_{ti} is the probability of system damage with respect to each group of elements of the same type which are included among tested devices; P_{nj} is the probability of system damage with respect to a group of analgous elements (elements of same type) which do not enter into the makeup of the tested devices; P_{sl} is the probability of system damage with respect to the remaining groups of elements which are not included among the tested devices and aggregates of the system; L is the number of different groups of elements of the same type in the first complex (which are included among the tested devices); R is the number of groups of elements which are of the same type as the elements in the first complex which are not included among the tested devices; S is the number of groups of elements which are not of the same type as the elements in the first and second complexes which are not included among the tested devices and blocks in the system.

We shall study the possibility of determining the individual terms of Eq. (4) when testing systems with an incomplete set of equipment.


Composition of the system
Elements in the tested devices
Elements in the devices which are not tested
Elements which are of the same type as those that are tested
Elements which are of the same type as those that are tested
1 2 3 . . . L
M1 M2 M3 . . . ML
1 2 . . . R
M1 M2 . . . MR
1 2 . . . S
M1 M2 . . . MS
Number of groups
Number of groups
Number of groups

The Probability of System Damage with Respect to Groups of Elements in the First Complex

The first complex of terms in Eq. (4) can be determined directly from the results of testing the devices in the system.

Assuming that the probabilities P_{ni} of system damage (with respect to each group of elements in the first complex) are approximately equal to the arithmetic mean of the relative number of faults (the frequence of failures) in the system with respect to the specified group of elements $\bar{E}_{ti} P_{ti} \approx \bar{E}_{ti} = \Sigma m / N$ in accordance with (2), we shall have

$$P_{c1} = \sum^{L} P_{t1} \approx \sum^{L} \overline{F}_{t1} = \sum^{L} \frac{\sum^{N} B_i}{N} = \sum^{L} \frac{\sum^{L} m_i}{N}, \quad (5)$$

where m_i is the number of faults in the system with respect to each group of elements in the first complex, N is the number of tested systems (groups of elements of the same type).

The probability of system damage with respect to the first complex of groups can also be expressed in terms of probability of the damage to the elements which enter into the composition of the groups in the first complex.

Substituting the value of P_{ti} into (5) in accordance with (3), we obtain

$$P_{t1} \approx \sum^{L} \sum^{M_i} P_e, \quad (6)$$

where M_i is the number of elements in each group, and L is the number of groups.

If the numerical characteristics of the reliability indices for the elements are detected with a mean square error σ_e during the process of testing the system, it follows that the mean square error in determining the reliability indices for the system (with respect to each group of elements in the first complex) will be determined by the following relationship in accordance with (3a)

$$\sigma_{ti} = \sqrt{\sum^{M_i} \sigma^2_{ei}} = \sigma_{ei} \sqrt{M_i}. \quad (7)$$

If the quantities σ_{ti} are determined with identical reliability, then the mean square deviation of the reliability indices of the system with respect to all groups of elements in the first complex is determined in accordance with (2a) by the formula

$$\sigma_{c1} = \sqrt{\sum^{L} \sigma^2_{ti}} = \sqrt{\sum^{L} M \sigma^2_{e1}} \quad (7a)$$

Expression (7a) demonstrates that the quantity σ_{ci} depends not only on σ_{ei} but also on the structure of the system, the number of elements M_i of the same type, and the number of groups L.

The System Damage Probability with Respect to Groups of Elements of the Second Complex

From the conditions governing the problem elements in the second complex are of the same type as the elements in the corresponding groups of the first complex.

If the operating conditions and the mode of operation for the elements in the untested devices and aggregates of the system do not differ substantially from the operating conditions and mode of the corresponding elements in the first complex, it follows that the results of testing elements in the first complex can be extended to apply to elements of the same type in the second complex.

In that case the second term of Eq. (4) can be determined indirectly by scaling the results of tests made on each group of elements in the first complex so that they apply to the corresponding elements in the second complex. Assuming that the damage probability for elements of the same type in the first and second complexes is approximately identical ($P_{e1} \approx P_{e2}$), we shall determine the system damager probability with respect to each group of elements in the second complex.

For a different number of elements M_1 and M_2 in groups of the same type in the first and second complexes this probability is determined by the relationship

$$P_{ni} = P_{ti} \frac{M_2}{M_1}. \quad (8)$$

The probability of system damage with respect to all groups of elements in the second complex is determined by analogy with (5) by the formula

$$P_{c2} = \sum^{R} P_{nj} = \sum^{R} P_{ti} \frac{M_2}{M_1}, \tag{9}$$

where M_{j1} and M_{j2} are the respective numbers of elements in groups of the same type in the first and second complexes. The expression for P_{c2} can be written in accordance with (6) in the form

$$P_{c2} \approx \sum^{R} \sum^{M_i} P_{ej} \tag{9a}$$

The mean square deviation of the reliability indices for the system with respect to each group of elements in the second complex is determined by analogy with (7) by the relationship

$$\sigma_{nj} = \sigma_{ej} \sqrt{M_{2j}}. \tag{10}$$

The mean square deviation of the reliability indices for the system with respect to all groups in the second complex isby analogy with (7a) equal to

$$\sigma_{c2} = \sqrt{\sum^{R} \sigma_{nj}^2} = \sqrt{\sum^{R} M_{2j} \sigma_e^2}. \tag{11}$$

Since the damage probability for elements of the second complex can be determined only indirectly in accordance with the conditions governing the problem (from the results of testing elements in the second complex), it follows that in order to determine the quantities δ_{nj} in Eq. (11) we can introduce an additional (methodological) error which is caused by the difference in the modes of operation for elements of the same type in the first and second complexes.

The magnitude of the resultant error will depend on the difference between the modes and on the relationship between the volumes (number of elements) M_{2j}/M_{1i} for groups of the same type in the first and second complexes.

For an insignificant difference in the modes of operation and a relatively small number of elements in the second complex the magnitude of this error can practically be neglected; it can be assumed that the damage probability of the system with respect to the second complex of groups can be determined with a sufficient degree of accuracy for practical purposes from the results of testing an incomplete set of equipment.

The System Damage Probability with Respect to Groups of Elements in the Third Complex

The probability of system damage with respect to groups of untested elements in the third complex is determined by the third term in Eq. (4). The magnitude of this term cannot be estimated from the results of testing the system, since elements in the third complex are not included among the tested devices and are not of the same type as the elements in the first and second complexes.

This term can be determined only analytically from the results of preliminary tests of the elements proper.

If as a result of testing these elements (under conditions corresponding to the operating conditions) the distribution of their faults $\varphi_i(t) \approx m_i(t)/M_{i3}$ is detected statistically along with the dispersion of this distribution, then it follows that on the basis of known postulates in probability theory it is possible analytically to determine the damage probability P_{e3} for these elements during the period corresponding to the testing time for the system.

In that case the system damage probabilities with respect to each group of elements in the third complex can be determined approximately in accordance with (3) by the relationship

$$P_{sl} \approx \sum^{M_l} R_{sl} \tag{12}$$

The probability of system damage with respect to all the groups of elements in the third complex is determined by analogy with (6) by the expression

$$P_{c3} \approx \sum^{S} P_{sl} = \sum^{S} \sum^{M_l} P_{esl}. \tag{13}$$

where M_l is the number of elements in each group, S is the number of groups of elements in the third complex.

The mean square deviation of the numerical characteristics for the system reliability indices with respect to all the groups of elements in the third complex is determined by analgy with (7a) by the equation

$$\sigma_{c3} = \sqrt{\sum^{S} \sigma_{sl}^2} = \sqrt{\sum^{S} M_{el} \sigma_{el}^2}. \tag{14}$$

The quantity σ_{c3} does not take into account the methodological error in estimating the reliability of the elements; this error may have an effect when there is a substantial difference between the testing and operating conditions for the elements in the thrid complex.

For an analytical estimate of system reliability it is assumed that the difference between the testing and operating conditions of these elements is not substantial, and thus the magnitude of the methodological error can be neglected since it is averaged.

The Error in Estimating the Reliability of Systems

If the probability of system damage with respect to all the complexes of element groups can be expressed in accordance with (4):

$$P_s \approx P_{c1} + P_{c2} + P_{c3},$$

then substituting the value of P_c we will have the approximate relationship

$$P_s \approx \sum^{L} P_{t_i} + \sum^{R} P_{n_j} + \sum^{S} P_{sl}$$

in accordance with (5), (9), and (13).

Replacing these system damage probabilities by the probability of element damage, we obtain the formula

$$P_s \approx \sum^{L} M_i P_{e_i} + \sum^{R} M_j P_{e_j} + \sum^{S} M_l P_{e_l},$$

where M_e and P_e are the number of elements of the same type in different groups of each complex and the probability of their damage during the time t.

The mean square deviation of the reliability indices for the system can in that case be expressed in terms of the mean square deviation of the terms

$$\sigma_s = \sqrt{\sigma_{c1}^2 + \sigma_{c2}^2 + \sigma_{c3}^2}. \tag{15}$$

Substituting the value of σ_{ci} from (7a), (11), and (14) into (15), we obtain

$$\sigma_s = \sqrt{\left(\sqrt{\sum^{L}\sigma_{ei}^2}\right)^2\left(\sqrt{\sum^{R}\sigma_{ej}^2}\right)^2 + \left(\sqrt{\sum^{S}\sigma_{el}^2}\right)}. \qquad (15a)$$

Replacing the terms in the right side with the mean square deviations of the reliability indices for the elements, (15a) can be written as

$$\sigma_s = \sqrt{\left\{\sqrt{\left[\sum^{L}\sqrt{M_i\sigma_{ei}^2}\right]^2}\right\}^2 + \left\{\sqrt{\left[\sum^{R}\sqrt{M_j\sigma_{ej}^2}\right]^2}\right\}^2 + \left\{\sqrt{\left[\sum^{S}\sqrt{M_l}\sigma_l^2\right]^2}\right\}^2}. \qquad (15b)$$

Expression (15b) demonstrates that the quantity σ_c depends not only on σ_{ei}, σ_{ej}, and σ_j, but also on the structure of the block diagram for the system: the number of groups containing elements of the same type L, R, and S, and the volumes of these groups (the number of elements in the groups M_i, M_j, and M_l)•.

For a system consisting of M_0 identical elements (moduli) that has a structure which is maximally uniform the value of the mean square deviation of the reliability indices for the system from their average probability value will in accordance with (15b) be equal to

$$\sigma_s = \sqrt{M_1\sigma_e^2 + M_2\sigma_e^2} = \sqrt{M_0\sigma_e^2},$$

where $M_0 = M_1 + M_2$ is the over-all number of elements in the system, M_1, M_2 are the number of elements in the tested and untested devices of the system.

It should be noted that irrespective of the degree of nonuniformity of the system structure it is possible to determine the over-all error more accurately in estimating system reliability for differentiated manipulation of the test results than for the conventional method of estimation; this is due to the additional averaging of the quantities σ_{ei}, σ_{ej} and σ_l (the conventional method involves an estimate over a set of systems).

The Error of the Reliability Estimate of Systems

If the error in estimating the mathematical expectation of the reliability indices for the elements can be expressed by the quantity

$$|a_e^* - a_e| = \gamma_e,$$

then the probability that the true value of the unknown quantity a_e will be confined within the limits $a_e^* - \gamma_e$; $a_e^* + \gamma_e$ is determined by the relationship••

$$\alpha_e = P(|a_e^* - a_e| < \gamma_e)$$

• The multiple nature of the averaging proves the possibility of determing the value of σ_s with a high accuracy.

•• In determining the statistical estimate of system reliability the quantities a_e^* are defined as the arithmetic mean of the relative number of faults (the frequency of failures) in the elements during the time t:

$$a_e^* = \frac{\sum^{M} n_{ei}}{M_i},$$

where n_{ei} is the number of failures for the i-th elements of the system during the time t, and M_i is the over-all number of (tested) elements of the same type.

The probability a_e in this case characterizes the reliability of estimating the quantity a_e, and γ_e represents the error in this estimate.*

It should be noted that a_e and γ_e are interrelated quantities since for a specified number of samplings M_i the reliability of estimating a_e increases to the extent that the allowable error γ_e increases; conversely, the decrease in the reliable interval $a_e^* \pm \gamma$ leads to a lowering of the reliability α.

For a specified reliability α the error γ in the estimate (for normal distribution of the deviations of the reliability indices for the elements) can be expressed in terms of the mean square error**.

If we use the quantity

$$\sigma_e^2 = \frac{1}{M-1}\sum^{M} (n_i - a_e^*)^2,$$

as an approximate value for the dispersion of the element failure frequency, then the approximate value for the dispersion of the mean arithmetic frequency of element failures is determined by the expression

$$\eta_e^2 = \frac{\sigma_e^2}{M} = \frac{1}{M(M-1)}\sum^{M} (n_i - a_e^*)^2.$$

In that case, as demonstrated in [1,2], for a specified reliability α the error in estimating the mathematical expectation of the reliability indices for the elements can be expressed using the relationship

$$\gamma_e = \eta_e \tau_\alpha, \tag{16}$$

where τ_α is a proportionality multiplier which is determined from tables cited in probability theory texts; this multiplier depends on the reliability of estimating α and the number of samplings M_i.

Relationship (16) makes it possible to determine the error in the statistical estimate of the reliability of the elements as a function of the number of tested elements M_i and the reliability α of the estimate.

If the quantities γ_{ei} are determined for elements in the system with an identical reliability γ, then by expressing γ_e in terms of σ_e we shall have by analogy with (15b)

$$\gamma_s = \sqrt{\left\{\sqrt{\left[\sum^{L}\sqrt{M_i\gamma_{ei}^2}\right]^2}\right\}^2 + \left\{\sqrt{\left[\sum^{R}\sqrt{M_j\gamma_{ej}^2}\right]^2}\right\}^2 + \left\{\sqrt{\left[\sum^{S}\sqrt{M_l\gamma_{el}^2}\right]^2}\right\}^2}. \tag{17}$$

Relationship (17) makes it possible to determine the experimental error in the statistical estimate of the reliability of systems on the basis of the results obtained by testing an incomplete set of equipment.***

It should be noted that as the number of elements M_i of the same type is increased the error in estimating the system reliability with respect to each group of elements (and therefore the error in estimating the reliability system as a whole) will tend toward zero regardless of the number of complexes in the tested equipment.

The validity of this proposition is based on the law of large numbers and can be verified using the Chebyshev inequality from which we obtain:

---

* Analogous concepts are also established for estimating another numerical characteristic of the reliability indices for the elements: the dispersion D_e or the mean square deviation $\sigma_e\sqrt{D_e}$. The accuracy and reliability of estimating the numerical characteristics of random quantities for a small number of samplings is treated in [1,2].

** It is assumed that when the technological process flows in a normal mode and the operating conditions are optimal the deviation of the reliability indices for the elements during fixed time intervals is approximately governed by a normal distribution. The validity of this proposition is demonstrated in [3].

*** We would have arrived at an analogous formula on the basis of an analytical estimate of system reliability on the basis of the specified reliability indices for its elements.

$$\lim P\{|a_e^* - a_e| \geqslant \gamma\} = 0 \quad (M \to \infty)$$

$$\lim P\{|a_e^* - a_e| \leqslant \gamma\} = 1 \quad (M \to \infty),$$

i.e., the probability of the deviation of the arithmetic mean for the random values of the element reliability indices a_e^* from the mathematical judgement a_e by an amount greater than $\gamma > 0$ tends toward zero for $M \to \infty$.

This proposition proves that the efficiency of the statistical estimate of system reliability using the differentiated method increases as the uniformity of the system structure increases.

Conclusion

The estimate of system reliability from the results obtained by testing an incomplete set of equipment can be performed using the differentiated method if statistical and indirect analytical methods for estimating the system errors from groups of elements of the same type, which are included in the tested and untested devices of the system, are combined.

The accuracy and reliability of such an estimate depend substantially on the uniformity of the system structure and on the relationship between the volumes of the element groups that make up the tested and untested devices.

LITERATURE CITED

1. E. S. Ventsel', Probability Theory [in Russian](State Technical Press, 1958)
2. B. R. Levin, Theory of Random Processes and Its Application in Radio-Engineering [in Russian] (Soviet Radio Press, 1957)
3. S. M. Kuznetsov, "The probability of damage to elements in automation systems", Avtomatika i Telemekhanika, 19, No. 11 (1958).

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

A FAST MULTICHANNEL DISTRIBUTOR USING TRANSISTORS

Ya. E. Belen'ky and V. N. Mikhailovsky (L'vov)
Translated from Avtomatika i Telemekhanika Vol. 22, No. 8,
pp. 1117-1122, August,1961
Original article submitted July 21, 1960

A multiphase multivibrator is described which is made up of transistors, using one transistor for each operation controlling pulse. Basic elements for computation of the circuit's operation are indicated, and possibilities for use of the scheme as a fast distributor are shown.

Distributors belong to the functional units most frequently used in distant control apparatus. They are employed for start-stop operations, for coding, as commutator devices, as selector switches, etc.

The basic goals in the development of distributors are improvements of accuracy, of dependability, of length of useful life, of speed of operation, and of flexibility in readjusting to the number of channels needed at various times, also the decrease of over-all dimensions of power consumption and of cost. These goals are reflected most highly in the distributor arrangements which are being assembled for articles of mass production.

Lately a most economical vacuum tube circuit for a distributor has been proposed [1]; this is the so-called multiphase multivibrator circuit, in which for each channel one vacuum tube and three passive circuit elements are required.

Circuits of the multiphase multivibrator type make it possible to obtain distributors with start-stop action, which have found widespread applications in distant control devices. Transistors used in the circuit as "active" components provide an ultimate simplification of the distributing device.

Let us examine the operation of a multiphase multivibrator using junction type p-n-p transistors under conditions of self-excitation. The circuit of such a self-excited transistorized multiphase multivibrator represents a multistage RC amplifier having a common emitter resistance r_e (Fig. 1) which in the general case can be replaced by a complex impedance Z_e.

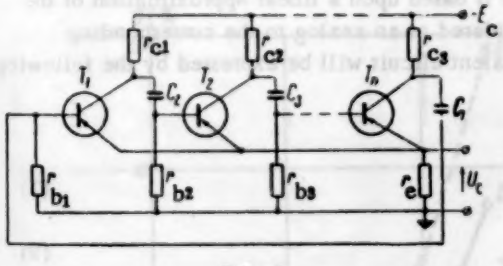


Fig. 1.

The resistances r_c and r_b and the capacitance C have the same values in all stages of the circuit ring chain of which a symmetrical multivibrator is composed. The workings of the steady state oscillatory regime of such a multivibrator can be represented as follows (Fig. 2).

At a certain moment of time all transistors of the multivibrator, with the exception of the first, are blocked due to the positive potential produced across resistance r_e. To transistor T_1, in addition, a negative potential bias U_m is applied, due to the preceding transient process which caused a change in the charge of condenser C_1 (Fig. 2a).

The potential U_{b1} exponentially decreases with time, and thereby causes an exponential falling off of the positive potential pulse in the collector line (Fig. 2b). Simultaneously, as a result of the decrease of the collector current i_c, the positive voltage bias U_e is decreased (Fig. 2d).

The positive impulse of the collector voltage with the exponential upper part is applied to the transition circuit loop $r_{b_2}C_2$ (which is a quasidifferentiating circuit), and then goes to the base of the next transistor T_2 (Fig. 2c). As a result of the quasidifferentiation of the exponential upper part of the collector impulse, the potential on r_{b2} becomes negative, and from then on during its further increase compensates the blocking action until finally transistor T_2 is opened up for passage of current.

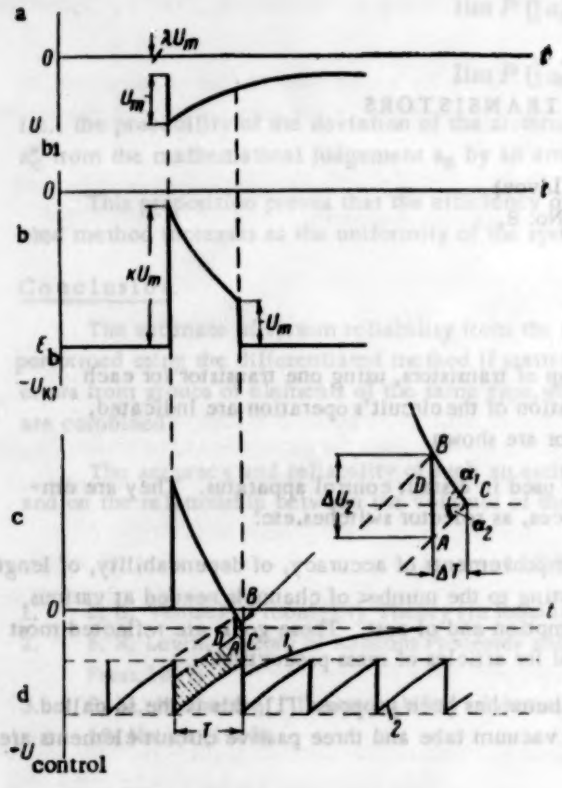


Fig. 2.

At that moment the processes in transistors T_1 and T_2 are similar to the processes in a multivibrator with common emitter at the moment of its flipping-over. The collector current of transistor T_2 produces an additional potential on resistance r_e, which decreases the collector current of the transistor, and this in turn produces an increase of the negative potential on r_{b2} and an increase of the collector current of transistor T_2, and so forth.

As a result of the flipping-over, the potential on the base of the second transistor is decreased by the amount of voltage change in the tail of the pulse in the collector circuit of transistor T_1. In the following course of events, the described process is repeated cyclically, proceeding along the transistors included in the following stages of the ring chain.

The transistors of a multivibrator ring chain, other than the two that take part in the generation of the particular pulse which is due to be produced, remain blocked at that time and have no effect on the over-all action of the entire circuit. Therefore, a multiphase multivibrator can be composed of a great number of stages.

If the ring chain of transistors is broken by applying to the base of the first transistor a negative potential bias large enough for the current of the first transistor to produce on the common emitter resistance a potential which will block all of the transistors, then the circuit will be in a stable condition.

A positive starter pulse, applied to the base of the first transistor, causes the blocking of this transistor and the unblocking of the next transistor in the chain - and then the generation of the successive pulses proceeds as described above; the total number of pulses produced is equal to the number of stages in the ring. After generation of the last impulse, the potential on the base will fall to zero, U_e will be decreased, the first transistor will open up, and the circuit will gain return to the stable state. In this way simulation of a standby multivibrator is obtained.

A mathematical analysis of the operation of such a multivibrator is based upon a linear approximation of the open circuit characteristics of a transistor. The transistor circuit is considered as an analog to the corresponding vacuum tube circuit. The gain and the internal resistance of the equivalent circuit will be expressed by the following formulas:

$$\mu_T = \frac{h_{21}}{\Delta h + r_e h_{22}}, \tag{1}$$

$$R_{iT} = \frac{h_{11} + r_e (1 + h_{21})}{\Delta h + r_e h_{22}}, \tag{2}$$

where the h_{ji}'s are the h parameters of the transistors with grounded emitters in matrix form, and

$$\Delta h = h_{11} h_{22} - h_{12} h_{21} + r_e h_{21}.$$

The duration of the generated pulses, with the assumptions mentioned above and under the condition that $r_c \ll r_b$, is determined by the following expression:

$$T = \tau \frac{k^2 a_2 (1 + a_1) + k [a_1 + a_2 (e - 1)] - 1}{k^2 a_2 + k a_2 (e - 2)}, \tag{3}$$

where

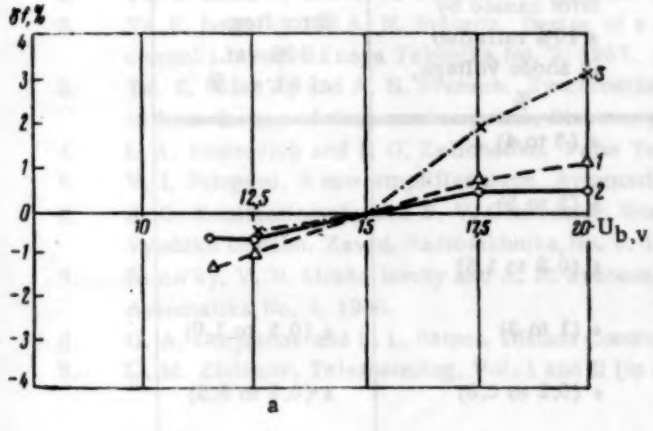
$$\tau = r_6 c,\ k = \frac{\mu_T}{1 + \frac{R_{iT} + r_D}{r_R}},\quad a_1 = e^{-\frac{1}{1+\lambda}},\quad a_2 = e^{-\frac{2}{1+\lambda}}.$$

The value of λ is found by means of experiments; it varies within the limits of from 0.2 to 0.5.

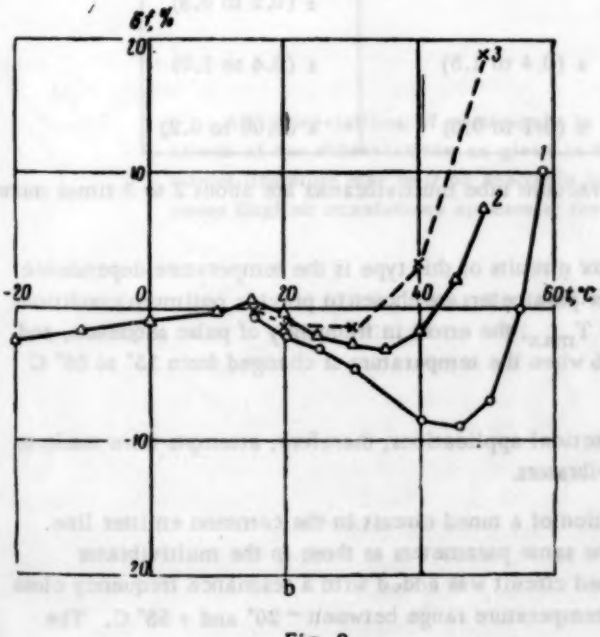
Practically the ratio of the duration T of the pulse to the time constant τ of the transfer circuit loop T/τ varies between 0.8 and 1.2 for various regimes of stable pulse generation.

It is evident from formula (3) that by changing the parameters of the RC circuit loops it is possible to change the duration of the diverse pulses, and thus it is possible to obtain a "nonsymmetrical" multiphase multivibrator with pulses of predetermined unequal durations [2].

The duration of the pulses can also be varied by changing the bias voltages applied to the bases of the various stages; by this means, an earlier or later unblocking of the transistor next in line can be obtained. This will make it possible to vary the output pulses in duration while the operation of the multivibrator is in progress.



a



b

Fig. 3.

The transistorized multiphase multivibrator circuit is well adapted to be synchronized with timing or starting pulses, and specifically on a standby basis; it also can be synchronized with sinusoidal and pulse voltages of external origin. The voltage to be synchronized with U_c is applied to the common branch of the emitter circuit (Fig. 1), and it is possible to specifically control the duration of each output pulse by means of an external voltage; yet only a single source of synchronization is needed. This makes it possible to use the multivibrator in various pulse techniques of distant control apparatus where certain definite time relationships are required between the appearances of the pulses from the distributor and the operation of other components.

A good example of application of the described methods is the synchronization of two commutators for multichannel measurements at a distance which was described in [3]. In the case there considered, the vacuum tubes are replaced by transistors.

When the problems of stability are examined, it is found that the sawtooth shape of voltage pulses in the emitter, as evident from Fig 2c (right side insert), increases the angle $\alpha_1 + \alpha_2$ made by the intersection of the base to ground voltage drop U_b (curve 1) with the emitter to ground voltage (curve 2); and the latter are the factors which control the unblocking of the next in line stage and determine the duration of the produced pulse.

The increase of the intersection angle of the controlling voltages, as is well known, lowers the lack of constancy of the pulse durations in the multiphase multivibrator circuit which is due to various factors: variations of supply voltage, aging temperature changes, etc.

An analysis shows that the advantage gained by the described circuit in this respect, in comparison with ordinary multivibrator circuits using junction triodes with

"negative" bases in which the steepness at the moment of unblocking of the transistors is at a maximum, reaches a value of 1.7 to 2 times. The most advantageous regime in a multivibrator is obtained when $T/\tau \approx 1$; this can be accomplished by variation of the ratio r_e/r_c.

In Fig. 3a are shown experimentally obtained relationships between the supply battery voltage U_b and the frequency variation percentage in a multiphase multivibrator using type P14 transistors. Curve 1 corresponds to a high ration of $r_e/r_c = 0.35$; curve 2 was taken at a ratio of $r_e/r_c = 0.3$ when the frequency reached a minimum ($r_b = 82T$; $r_c = 20T$; $r_e = 6T$); and curve 3 corresponds to a regime of dc currents in the base to emitter connection. The experimental results show that maximum stability is obtained when $T = T_{max}$, and that at $E_e = -15v$ the frequency varies not more than 0.3 to 0.7% when the supply voltage changes by + 10%.

The highest percentage variation of frequency δf (obtained with various parameters in the multivibrator circuit when stable conditions of excitation are maintained) did not exceed 2 to 3% for supply voltage variations of ± 10%.

Experimental data for multivibrator circuits using vacuum tubes have shown that their instability is less than ± 0.2% for ± 10% changes in plate and filament voltage.

All data from experiments on transistor and vacuum tube multivibrators are collected in the following table; the data for vacuum tube multivibrators are taken from various publications [4 to 6]. Included are also some data on known single phase multivibrators.

Type of multivibrator	Error caused by ± 10% variation in anode voltage, %	Error from ± 10% var. of fil. v., %
Semiconductor multivibrator [6]	± (3 to 4)	
Multiphase semiconductor multivibrator (maximum stray)	± (2 to 3)	
Multiphase semiconductor multivibrator at $T \approx T_{max}$	± (0.2 to 1.5)	
Selfexcited vacuum tube multivibrator with zero grid bias [4,5]	± (1 to 3)	± (0.5 to 1.0)
Vacuum tube multivibrator with "positive" grid [4]	± (0.2 to 0.5)	± (0.2 to 0.5)
Vacuum tube multivibrator with cathode coupling [4]		± (0.2 to 0.3)
Multiphase multivibrators using vacuum tubes (maximum stray)	± (0.4 to 1.5)	± (0.4 to 1.0)
Multiphase multivibrators using vacuum tubes with $T \approx T_{max}$	± (0.1 to 0.3)	± (0.06 to 0.2)

The table shows that the multiphase semiconductor and vacuum tube multivibrators are about 2 to 3 times more stable than known two phase multivibrators.

Among the most sensitive characteristics of semiconductor circuits of this type is the temperature dependence of the duration of the generated pulses. When the multivibrator parameters are chosen to provide optimum conditions with respect to constancy of operation; specifically, when $T \approx T_{max}$, the errors in frequency of pulse sequence, and consequently also the errors in pulse durations, reach about 10% when the temperature is changed from 15° to 55° C (Fig. 3b, curve 1).

Obviously, this can not always satisfy requirements of practical applications; therefore, attempts were made to substantially improve the temperature characteristics of multivibrators.

A positive result in this respect was obtained by introduction of a tuned circuit in the common emitter line. Curve 2 on Fig. 3b was taken with a multivibrator having all the same parameters as those in the multivibrator used for curve 1, except that in the common emitter line a tuned circuit was added with a resonance frequency close to the average frequency of the sawtooth voltage pulses in the temperature range between − 20° and + 55° C. The

instability (range of fluctuations) of sequence frequency of the multivibrator in the temperature range between 15° and 55°C was changed by this device 3 to 5 times. With different parameters of the multiphase multivibrator circuit which do not provide optimum conditions, the presence of the tuned circuit also increases the frequency stability with respect to effects of temperature, but only up to about 40°C (curve 3 on Fig. 3b). If such a stabilization would have to be carried out for an ordinary multivibrator circuit of n stages, there would have to be used n such tuned circuits.

Introduction of reactive elements into the common channel and into the individual stages will make it possible to shape various functions at the moments of generation of the pulses in the individual stages, and to obtain sequences of functional curves shifted in time or space.

Thus, a multiphase multivibrator can be used successfully not only when it is made up with vacuum tubes, but also when it uses transistors.

Considering the possibilities offered by the described distributors, it is not difficult to visualize a number of applications where they can be used to advantage, such as in commutators on the sending and receiving stations of multichannel telemetering systems [7], in input units of coding devices [8], as start-stop devices [8], in fast acting distributors of digital and pulse coding distant control systems [8,9], and also in all other pulse devices where sequences of pulses are required which are shifted in time by intervals of specified durations.

LITERATURE CITED

1. Ya. E. Belen'ky and A. N. Svenson. Multiphase multivibrator, Radiotekhnika, No. 7, 1956.
2. Ya. E. Belen'ky and A. N. Svenson. Design of a nonsymmetrical multiphase multivibrator, Avtomaticheskii control i Izmeritel'naya Tekhnika, No. 2, 1957.
3. Ya. E. Belen'ky and A. N. Svenson. Synchronization of commutators in multichannel communication systems without the use of time marker pulses, Elektrosvyaz', No. 12, 1957.
4. L. A. Meerovich and L. G. Zelichenko. Pulse Technics [in Russian], Soviet Radio Press, 1953.
5. V. I. Pampuro, A new amplifier stage, Avtomatika, No. 1 (1958).
6. A. G. Konstantinovsky and V. V. Chervetsov, Stabilization of multivibrators using transistors, Izvestiya Vysshikh Uchebn. Zaved. Radiotekhnika, No. 5, 1958.
7. Belen'ky, V. N. Mikhailovsky and A. N. Svenson. Solution of circuit problem in multichannel commutation, Avtomatika, No. 4, 1956.
8. O. A. Goryainov and R. L. Raines. Distant Control [in Russian] State Power Press, 1954.
9. C. M. Zhdanov. Telemetering, Vol. I and II [in Russian] (State Power Press, 1952).

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

THE STABILITY OF SINGLE-STAGE APERIODIC LOW-FREQUENCY VACUUM TUBE AMPLIFIERS

V. I. Pampuro (Kiev)

Translated from Avtomatika i Telemekhanika Vol. 22, No. 8,
pp 1123-1132, August, 1961
Original article submitted July 28, 1960

The paper studies the stability of aperiodic amplifiers over a small time interval (6 to 8 hours) during which the variation of the element parameter in the circuit is much smaller in absolute magnitude than the modulus of the parameter. The concept of the instability coefficient is introduced; this coefficient is determined for a number of practical circuits. The boundary tests are also studied.

In this paper we study the stability of single-stage aperiodic amplifiers [1]. The stability of the parameters is studied over a small time interval (6 to 8 hours); this has meaning for instrumentation amplifiers which are calibrated before each set of measurements. For this limitation the modulus of the increment in the element parameters of the circuits is much smaller than the modulus for the nominal parameter of the circuit element.

An analysis of the stability of the circuit parameters when the element parameters vary is performed according to the Polivanov formula generalized by Sigorskii [2]. For example, the gain of an amplifier which is represented as a fourpole with a short-circuited side having an input terminal a and an output terminal b can be written as

$$\dot{K} = \frac{\Delta_{ab}}{\Delta_{aa}}, \tag{1}$$

where Δ_{ab} and Δ_{aa} are the algebraic additions to the principal determinant Δ of the admittance matrix. Its increment [2] is

$$\delta\dot{K} = \frac{\Delta^{\circ}_{ab}\Delta_{aa} - \Delta^{\circ}_{aa}\Delta_{ab}}{\Delta_{aa}(\Delta_{aa} + \Delta^{\circ}_{aa})}, \tag{2}$$

where Δ°_{ab} and Δ°_{aa} are the increments of the corresponding determinants when the element parameters are varied (cf. Appendix II). Since the absolute magnitude of the variation of the element parameters is much less than the parameter modulus $\Delta_{aa} \gg \Delta^{\circ}_{aa}$, it follows that

$$\delta\dot{K} = \frac{\Delta^{\circ}_{ab} - \dot{K}\Delta^{\circ}_{aa}}{\Delta_{aa}}. \tag{3}$$

The amplifier stage under study does not include the output admittance Y_g of the input generator; i.e., the gain $\dot{K}$ determines the ratio between the voltage U_b at the output and the voltage U_a at the input. For the majority of circuits analyzed in this paper this leads to a situation where the internal feedback through the interelectrode capacitances is not taken into account. Therefore, it is necessary to define the gain $\dot{K}_E$ as the ratio between the voltage at the output of the circuit and the generator emf. The gain $\dot{K}_E$ can be written expediently if we express it in terms of the amplifier input admittance Y_{in}, the admittance Y_g, and the gain $\dot{K}$:

$$\dot{K}_E = \frac{Y_g}{Y_g + Y_{in}}\dot{K}. \tag{4}$$

From the formula cited above it is evident that the gain $\dot{K}_E$ consists of two parts: the first $Y_g/(Y_g+Y_{in})$ is the variable part which depends on the generator connected to the input of the circuit (the internal admittance of the generator differs for each specific case); the second part is the coefficient K which is constant for the specified circuit. It is evident that we can treat the coefficient $\dot{K}$ for a specified circuit as a quantity that will apply in all cases; the first portion is analyzed separately in each case. Moreover, the input admittance Y_{in} can be expressed in terms of the coefficient $\dot{K}$, and thus its variation can be estimated in terms of the increment in the coefficient $\dot{K}$.

The increment of the coefficient δK can be determined by neglecting quantities of second order.

The stabilizing qualities of the circuit are determined in terms of the instability coefficient H, which shall be assumed to denote the ratio between the relative variation of the circuit parameter for variation of the element parameters and the square root of the quotient obtained by dividing the sum of the squares of the relative parameter variations X_i by the number of elements $\underline{n}$. For example, the instability coefficient with respect to the modulus K of the gain is

$$HK = \frac{\delta K / K}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\delta X_i / X_i)^2}}. \qquad (5)$$

A numerical determination of the relative variation of a circuit parameter depends on whether or not the element parameters are statistically interrelated. For example, in determining the over-all increment in the transconductance of a two-tube circuit it is possible to assume that the variations are statistically interrelated if we investigate the effect of the voltage of a common supply source; we can assume that they are not statistically interrelated if we investigate the effect of aging or other factors which are difficult to consider exactly.

As we shall demonstrate below, the approximate expression for the gain δK_0 can be expressed in terms of two increments of element parameters that in the general case are not satistically interrelated. A determination of the instability coefficient according to formula (5) leads to an appreciable error. Therefore, we shall determine the maximum possible error; i.e., we shall determine the maximum instability coefficient with respect to modulus in accordance with the formula

$$HK_{\max} = 2\frac{\left|\frac{\delta K_{0\max}}{K_0}\right|}{\left|\frac{\delta X_1}{X_1}\right| + \left|\frac{\delta X_2}{X_2}\right|}, \qquad (6)$$

where K_0 is the gain at intermediate frequencies.

Cases of statistically interrelated parameters shall be studied separately in analyzing each circuit.

The analysis of the circuits is achieved by the method of node voltages [2]. The results of this analysis are cited in the table.

The complete expression for the increment in the coefficient $\delta \dot{K}$ for variation of element parameters takes into account the variation of both the resistive and the reactive elements. The parameters of the reactive elements and of certain of the resistive elements (for example, the admittance of the grid circuit) affect the gain basically at the boundaries of the frequency response. Therefore, in a first approximation their variation can be neglected, and this makes it possible for us to obtain a simplified expression for the gain at intermediate frequencies.

In those cases when a triode is used, we use the substitution $\delta S = \mu \delta G_i$ in deriving the approximate expressions for the coefficient δK_0 ; here S is the transconductance of the tube, μ is the amplification factor, and G_i is the internal conductance of the tube. Here it is assumed that the amplification factor is constant with time when the external conditions change. The constant nature of the coefficient μ is explained on the basis of the fact that during a small time interval the coefficient μ will be affected only by a variation in the external conditions; this variation consists of a variation of the supply voltages, and it is known that the coefficient μ remains constant when this variation is small [3].

TABLE

Circuit		Fig. 1	Fig. 4
$\dot{K}$	Triode	$-\frac{S-Y_3}{Y_1+Y_3+Y_4}$	$\frac{S+Y_1+Y_3}{Y_1+Y_3+Y_4}$
	Pentode	$-\frac{S-Y_3}{Y_3+Y_4}$	
K_0	Triode	$-\frac{S}{G_1+G_4}$	for $S \gg G_1$ $\frac{S}{G_1+G_4}$
	Pentode	$-\frac{S}{G_4}$	
$\delta\dot{K}$	Triode	$-\frac{(\delta S-\delta Y_3)+\dot{K}(\delta Y_3+\delta Y_4+\delta Y_1)}{Y_1+Y_3+Y_4}$	$\frac{(\delta S+\delta Y_1+\delta Y_3)-\dot{K}(\delta Y_1+\delta Y_3+\delta Y_4)}{Y_1+Y_3+Y_4}$
	Pentode	$-\frac{(\delta S-\delta Y_3)+\dot{K}(\delta Y_3+\delta Y_4)}{Y_3-Y_4}$	
δK max	Triode	$K_0\frac{G_4}{G_1+G_4}\left(\left\lvert\frac{\delta G_1}{G_1}\right\rvert+\left\lvert\frac{\delta G_4}{G_4}\right\rvert\right)$	$K_0\frac{G_4}{G_1+G_4}\left(\left\lvert\frac{\delta G_1}{G_1}\right\rvert+\left\lvert\frac{\delta G_4}{G_4}\right\rvert\right)$
	Pentode	$K_0\left(\left\lvert\frac{\delta S}{S}\right\rvert+\left\lvert\frac{\delta G_4}{G_4}\right\rvert\right)$	
$\dot{K}$ max	Triode	$2\frac{G_4}{G_1+G_4}$	$2\frac{G_4}{G_1+G_4}$
	Pentode	2	
Quantity $\dot{K}$ max	Triode	for $G_1=3G_4$ 0,5	for $G_1=3G_4$ 0,5
	Pentode	2	
Y_{in}		$Y_3+Y_3(1-\dot{K})$	$Y_3+Y_3(1-\dot{K})$

The first column in the table shows the results of analyzing circuits with a common cathode (Fig. 1). From the expression for the increment δK_0 it follows that its magnitude decreases as the plate conductance G_4 decreases (for a triode). This can be explained on the basis of the fact that when the conductance G_4 decreases the coefficient K_0 tends toward μ and is practically independent of variations in the magnitude of the conductance G_1 when the condition $G_4 \ll G_1$ holds. However, it is difficult to insure the fulfillment of such a condition, since it requires a high-voltage supply source. For example, in order to insure a coefficient of instability with respect to modulus equal to 2% it is necessary to satisfy the condition $G_1 = 99\ G_4$; here the supply source must produce a voltage 100 times the plate voltage. Usually the plate load conductance G_4 is three to four times smaller than the conductance G_1 (for a triode). The pentode circuit operates in a constant current mode ($G_1 \ll G_4$) and is more sensitive to a variation of the parameters of the circuit elements (the tube parameters). Therefore, in practical circuits the stability of a pentode circuit is three to four times lower than that of a triode.

Continuation of table

Circuit		Fig. 3	Fig. 2
K	Triode	$\dfrac{S+Y_3}{Y_1+Y_3+Y_4+S}$	$\dfrac{(S_1-Y_3)(S_2+Y_{i2})}{Y_4(S_2+Y_{i2})+(Y_{i1}+Y_3)(Y_{i2}+Y_4)}$
	Pentode	$\dfrac{S+Y_3}{Y_3+Y_4+S}$	
K_0	Triode	$\dfrac{S}{G_1+G_4+S}$	$-\dfrac{S_1}{G_4}$
	Pentode	$\dfrac{S}{G_4+S}$	
δK	Triode	$\dfrac{(\delta S+\delta Y_3)-K(\delta Y_1+\delta Y_3+\delta Y_4+\delta S)}{Y_3+Y_4+Y_1+S}$	$-((S_1-Y_3)(\delta S_2-\delta Y_{i2})+(\delta S_1-\delta Y_3)\times$ $\times(S_2+Y_{i2})+K[Y_4(\delta S_2+\delta Y_{i2})+$ $+(Y_{i1}+Y_3)(\delta Y_{i2}+\delta Y_4)+\delta Y_4(S_2+Y_{i2})+$ $+(\delta Y_{i1}+\delta Y_3)\times$ $\times(Y_{i2}+Y_{i4})]):[Y_4(S_2+Y_{i2})+$ $+(Y_{i1}+Y_3)(Y_{i2}+Y_4)]$
	Pentode	$\dfrac{(\delta S+\delta Y_3)-K(\delta Y_3+\delta Y_4+\delta S)}{Y_3+Y_4+S}$	
δK max	Triode	$K_0\dfrac{G_4}{G_1+G_4+S}\left(\left\lvert\dfrac{\delta G_1}{G_1}\right\rvert+\left\lvert\dfrac{\delta G_4}{G_4}\right\rvert\right)$	$K_0\left(\left\lvert\dfrac{\delta S_1}{S_1}\right\rvert+\left\lvert\dfrac{\delta G_4}{G_4}\right\rvert\right)$
	Pentode	$K_0\dfrac{G_4}{G_4+S}\left(\left\lvert\dfrac{\delta S}{S}\right\rvert+\left\lvert\dfrac{\delta G_4}{G_{14}}\right\rvert\right)$	
HK max	Triode	$2\dfrac{G_4}{G_1+G_4+S}$	2
	Pentode	$2\dfrac{G_4}{G_4+S}$	
Quantity HK max	Triode	for $S=99\,G_4$, $G_1\geqslant 3G_4$ 0,019	2
	Pentode	for $S=99G_4$ 0,02	
Y_{in}		$Y_3+\dfrac{Y_3}{1-\dot{K}'}$; $\dot{K}'=\dfrac{-S-Y_3}{Y_1+Y_4}$	$Y_3+Y_3(1-\dot{K}_{23})$ $\dot{K}_{23}=\dfrac{\dot{K}}{\mu_3}\left(1+\dfrac{Y_{i1}}{Y_{i2}}\right)$

The modulus instability coefficient for a pentode circuit is equal to unity. This means that a variation in the transconductance S or the conductance G_4 leads to a variation in the coefficient K, which is the same percentage-wise.

An analysis of the cascode circuit with a plate resistor (Fig. 2) leads to the same results as the analysis of a pentode circuit with a common cathode. This can be explained on the basis of the fact that the parameters - the equivalent transconductance S_e and G_{ie} - are given by the relationships*

$$S=S_1\frac{S_2+G_{i2}}{G_{i1}+G_{i2}+S_2},\qquad G_i=\left(1-\frac{S_2+G_{i2}}{G_{i1}+G_{i2}+S_2}\right)G_{i2}$$

Continuation of table

Circuit	Fig. 5	Fig. 6
$\dot{K}$	$-\frac{S_1+S_2-Y_3'+Y_3''}{Y_{i1}+Y_{i2}+Y_3'+Y_4}$	$-\frac{S_1-Y_3}{Y_{i1}+Y_{i2}+Y_3+Y_4}$
K_0	$-\frac{S_1+S_2}{G_{i1}+G_{i2}}$	$-\frac{S_1}{G_{i1}+G_{i2}}$
$\delta\dot{K}$	$-\frac{(\delta S_1+\delta S_2-\delta Y_3'+\delta Y_3'')}{Y_{i1}+Y_{i2}Y_3'+Y_4}-$ $-\frac{\dot{K}(\delta Y_{i1}+\delta Y_{i2}+\delta Y_3'+\delta Y_4)}{Y_{i1}+Y_{i2}+Y_3'+Y_4}$	$-\frac{(\delta S_1-\delta Y_3)+\dot{K}(\delta Y_{i1}+\delta Y_{i2}+\delta Y_3+\delta Y_4)}{Y_{i1}+Y_{i2}+Y_3+Y_4}$
δK_{max}	$\frac{S_1G_{i2}-S_2G_{i1}}{(G_{i1}+G_{i2})^2}\left(\left\lvert\frac{\delta G_{i1}}{G_{i1}}\right\rvert+\left\lvert\frac{\delta G_{i2}}{G_{i2}}\right\rvert\right)$	$K_0\frac{G_{i2}}{G_{i1}+G_{i2}}\left(\left\lvert\frac{\delta G_{i1}}{G_{i1}}\right\rvert+\left\lvert\frac{\delta G_{i2}}{G_{i2}}\right\rvert\right)$
HK_{max}	$2\frac{S_2G_{i1}-S_1G_{i2}}{(S_1+S_2)(G_{i1}+G_{i2})}$	$2\frac{G_{i2}}{G_{i1}+G_{i2}}$
Quantity HK_{max}	for $S_1G_{i2}=0{,}9S_2G_{i1}$ $0{,}05$	for $G_{i1}=10\div100G_{i2}$ $0{,}2\div0{,}05$
Y		$Y_2+Y_3(1-\dot{K})$

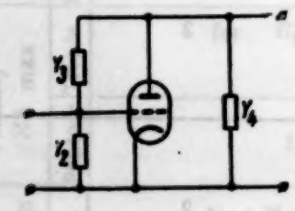


Fig. 1.

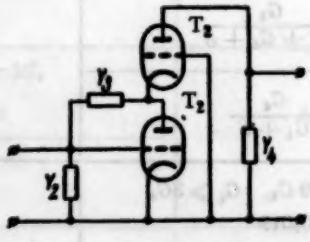


Fig. 2.

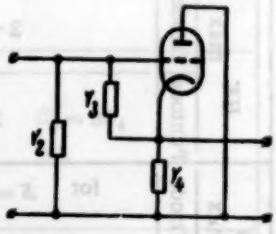


Fig. 3.

and are close to the pentode parameters (here S_1, G_{i1} and S_2, G_{i2} are the transconductance and conductance of the lower and upper tubes, respectively).

For resistive coupling of the grid of the upper tube T_2 to the cathode through a rheostat divider the plate voltage of the lower tube T_1 on which the gain depends is stabilized due to the effect of the current feedback introduced by the upper tube T_2. This factor and the absence of a screen grid lead to a situation where the cascode circuit is more stable than a pentode circuit with a common cathode.

The cathode follower circuit (Fig. 3) has a high stability. The stability of the circuit increases as the gain approaches unity. In the general case the stability of a cathode follower is determined by the voltage transmission over the feedback loop. A pentode circuit produces the maximum voltage transmission over the feedback loop. However, the transfer coefficient over the feedback loop cannot be increased without limit, since for a capacitive load the circuit can be excited [4]. The results cited in the table for a cathode follower are valid when its input circuit; the voltage from the output of the circuit is divided between the admittances Y_g and Y_{in}; under these conditions the amount of feedback is reduced and the stability drops.

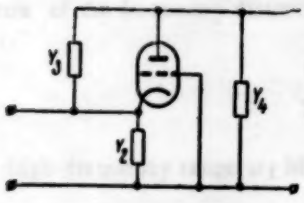


Fig. 4.

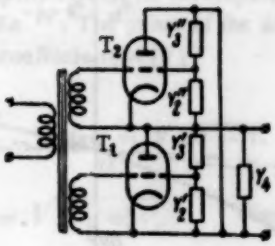


Fig. 5.

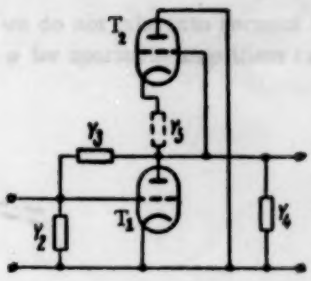


Fig. 6.

Figure 4 shows a circuit with a common grid. For an increment of the coefficient δK_0 in this circuit, we arrive at the same conclusions as those which apply to a circuit with a common cathode. The stabilizing effect of the current feedback can be taken into account in the coefficient K_E.

The cascode circuit with an input transformer (Fig. 5) [5] was analyzed without taking the transformer ratio into account. The computing formulas are cited in the table. The high stability of the circuit is explained by its balancing properties: when the supply voltage varies the variations of the circuit-element parameters are statistically interrelated. For small variations of the supply voltage the amplification factor μ remains constant, and the variations of the internal admittances of the tube are approximately equal in magnitude and sign; this leads to high stability for the circuit.

The pronounced balancing properties of the circuit lead to high stability for statistically non-interrelated element parameters as well.

A circuit with a tube in the plate circuit (Fig. 6) also has pronounced balancing properties. If the parameters of the circuit elements are not statistically interrelated, then the specified circuit is formally equivalent to a circuit with a common cathode. However, it may insure a small equivalent admittance in the plate of the first tube as compared to the admittance of the tube T_1 (for this purpose we connect the admittance Y_s) for a relatively small source-voltage supply; therefore, it has advantages over the circuit shown in Fig. 1. A shortcoming of the circuit is its high sensitivity to phase shifts between the current and the output voltage; this leads to a deterioration of the stabilizing properties of the circuit, and the frequency increases.

The analysis which we have made of circuit stability does not take into account the effect of the admittance Y_g. Usually it is through this admittance that parallel negative feedback is achieved due to parasitic capacitance between the output and input (for the circuits shown in Figs. 1,4,6); this feedback increases the input admittance of the circuit and therefore raises the instability associated with the variation in input admittance. The methods for reducing the input admittance consist of reducing the parasitic admittance between the output and the input (for a pentode), or of reducing the voltage transfer coefficient at the point at which the admittance is connected (parasitic capacitance). The latter occurs in a cascode circuit where an effort is made to reduce the coefficient K_{23} to a minimum (i.e., to reduce the coefficient for voltage transfer from node 2 to node 3 to a minimum).

The cathode follower has a very low input admittance and is usually employed as an impedance transformer. However, it should be said that this property of a cathode follower loses its value when the generator admittance Y_g is comparable to the input admittance; here the amount of feedback depends on the stability of the admittance and the stability of the cathode follower drops.

Thus, the cathode follower circuit, the pentode common-cathode circuit, and the cascode circuit are characterized by the lowest sensitivity to a variation in the admittance Y_g.

The verification of the circuit stability is associated with boundary tests [6] which presuppose determination of stability for specified variations of the element parameters. One of the simplest methods for varying the element parameters is the variation of the supply voltages; in particular, it is most convenient to vary the heater voltage. The boundary tests include variation of the stability of the circuit parameters, which is related to the tolerances of the components. Since the expression for the instability coefficient consists of the relative magnitudes of the element-parameter variations, it follows that the expression for the coefficient also takes into account the parameter spread due to tolerances or other factors that cause a variation in the element parameters. The variation of the instability

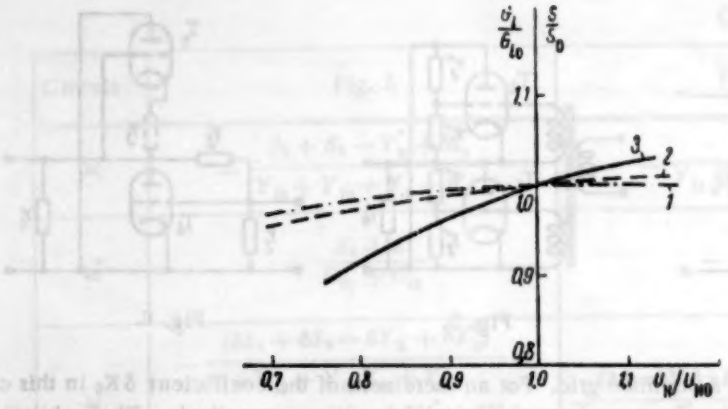


Fig. 7.

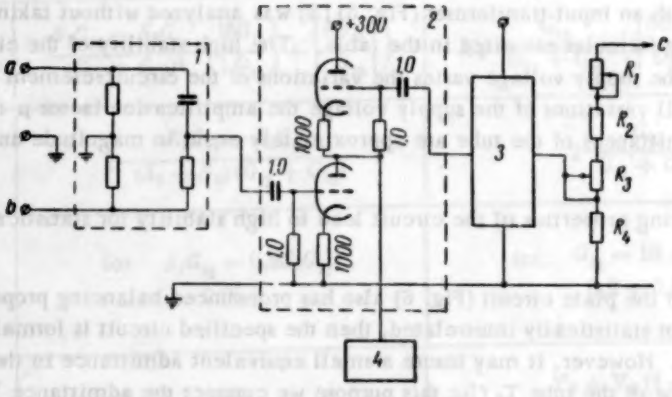


Fig. 8.

coefficient due to aging of the parts is determined by boundary tests which can expediently be combined with the calibration process. For a quantitative determination of the relationship between the instability coefficient and the element-parameter variation Fig. 7 shows the dependence of the transconductance S and the conductivity G_i on the variation of the heater voltage U_H(the curves $\frac{S}{S_0} = f\left(\frac{U_H}{U_{H0}}\right)$ and $\frac{G_i}{G_{i0}} = f\left(\frac{U_H}{U_{H0}}\right)$ coincide, since μ = const.) The number 1 denotes the curve $\frac{S}{S_0} = f\left(\frac{U_H}{U_{H0}}\right)$ for a "6N1P" tube; the number 2 denotes the corresponding curve for a "6N2P" tube, and the number 3 denotes the corresponding curve for a "6N14P" tube.

In practice the verification of the stability of single-stage aperiodic vacuum tube amplifiers was performed using the circuit shown in Fig. 8. The circuit consists of a phase shifter 1, a subtracting circuit 2, an investigated amplifier 3, and a vacuum voltmeter 4. The signal from the generator is applied at the points a, b; point c is connected to points a or b, depending on the number of amplifier stages. Having chosen the voltage at the amplifier input and the generator frequency, we balance the subtracting circuit for a minimum voltmeter reading (the circuit permits us to obtain a subtraction error no greater than 0.5% for an input voltage (1b) to the subtracting circuit). Having changed the heater voltage of the amplifier, we shall determine the voltmeter voltage 4; the ratio between the voltage which we obtain and the input voltage (1b) determines the relative variation of the gain for the amplifier 3. The experimental and computed (tabulated) data are in good agreement.

In this paper we have analyzed voltage amplifiers, and therefore the load of the amplifier output is not taken into account

APPENDIX I

Our analysis of the stability of the amplifiers will be incomplete if we do not take into account the instability with respect to the phase of the gain $\dot{K} = -Ke^{-j\varphi}$. The phase of the angle φ for aperiodic amplifiers can be determined in terms of the frequency distortion coefficient M [7]

$$\operatorname{tg}\varphi = \sqrt{M^2 - 1}.$$

In the high-frequency range $\omega_t M_t = \sqrt{1 + \omega_t^2 \tau_t^2}$. Then

$$\operatorname{tg}\varphi_t = \omega_t \tau.$$

Here τ_t is a time constant which is equal to the product of the normalized resistance and capacitance $R_t C_t$, which are determined from the denominator of the coefficient $\dot{K}$ (cf. Table) for each circuit.

For low frequency distortion the phase angle is equal to

$$\varphi_t = \omega_t \tau_t = \omega_t R_t C_t.$$

whence the maximum phase instability coefficient is

$$H\varphi_{t\max} = 2\frac{\left|\frac{\delta\varphi_t}{\varphi_t}\right|}{\left|\frac{\delta R_t}{R_t}\right| + \left|\frac{\delta C_t}{C_t}\right|} = 2.$$

An analysis at low frequencies leads to an analogous result. The phase instability coefficient for the analyzed aperiodic stages is identical.

APPENDIX II

Assume Δ is the determinant for the admittance matrix before variation of the parameters

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}. \tag{II.1}$$

Since the subscripts of an element in the determinant correspond to the numbers of the row and column whose intersection determines the position of the element, it follows that henceforth we shall introduce a coordinate dotted grid which eliminates the subscripts. Here in the general case the elements of different coordinate rows or columns are not equal to one another:

$$\Delta = \begin{array}{c|cccc} & 1 & 2 & \dots & n \\ 1 & a & a & & a \\ 2 & a & a & & a \\ \dots & \dots & \dots & \dots & \dots \\ n & a & a & & a \end{array}. \tag{II.2}$$

In view of the variation of the parameters of the determinant elements the determinant changes its value. Each element will consist of the quantity a and an increment b:

$$\Delta' = \begin{array}{c|cccc} & 1 & 2 & \dots & n \\ 1 & a+b & a+b & \dots & a+b \\ 2 & a+b & a+b & \dots & a+b \\ \dots & \dots & \dots & \dots & \dots \\ n & a+b & a+b & \dots & a+b \end{array}. \tag{II.3}$$

The theorem for expanding the determinant into a sum of the determinants Δ^{*} of the same order. A determinant Δ^{*} of n-th order, in which each element represents the sum of several terms a and b, can be expanded into a sum of determinants of the same order.

$$\Delta^{*} = \Delta + \Delta^{\circ}, \tag{II.4}$$

where the determinant Δ° is equal to

$$\Delta^{\circ} = \sum_{i=0}^{n-1} P_{i(n-i)}\ \Delta_{a_i b_{n-i}}, \tag{II.5}$$

$\Delta a_i b_{n-i}$ are n-th order determinants which are obtained by transposing the terms in the elements a and be for the case where the elements a and b repeat i and (n-i) times, respectively. $P_{i(n-1)}\ \Delta_{a_i b_{n-1}}$ denotes the sum of the determinants (after the tranpositions have been made) for the condition i = const. The number of determinants which enter into that sum is equal to $\frac{n!}{i!\,(n-i)!}$. The over-all number of determinants (5) is computed from the formula:

$$T = \sum_{i=0}^{n-1} \frac{n!}{i!\,(n-i)!}, \tag{II.6}$$

where in the case i = 0 we assume T = 1 (under these conditions the determinant consists exclusively of b-terms). For example, for a third order determinant T = 7; the determinant

$$\Delta^{\circ} = \begin{vmatrix} 1 & 2 & 3 \\ 1 & a & b & b \\ 2 & a & b & b \\ 3 & a & b & b \end{vmatrix} + \begin{vmatrix} 1 & 2 & 3 \\ 1 & b & a & b \\ 2 & b & a & b \\ 3 & b & a & b \end{vmatrix} + \begin{vmatrix} 1 & 2 & 3 \\ 1 & b & b & a \\ 2 & b & b & a \\ 3 & b & b & a \end{vmatrix} + \begin{vmatrix} 1 & 2 & 3 \\ 1 & a & a & b \\ 2 & a & a & b \\ 3 & a & a & b \end{vmatrix} + \begin{vmatrix} 1 & 2 & 3 \\ 1 & a & b & a \\ 2 & a & b & a \\ 3 & a & b & a \end{vmatrix}$$
$$+ \begin{vmatrix} 1 & 2 & 3 \\ 1 & b & a & a \\ 2 & b & a & a \\ 3 & b & a & a \end{vmatrix} + \begin{vmatrix} 1 & 2 & 3 \\ 1 & b & b & b \\ 2 & b & b & b \\ 3 & b & b & b \end{vmatrix}.$$

We shall assume that the term b is an increment of the element a or a complex component of it, or any other component which it is desirable to isolate.

LITERATURE CITED

1. G. A. Shevtsov, Analysis of Sensitivity and Errors of Electronic Instrumentation Amplifiers, Dissertation For A Candidate of Sciences Degree, Lwow Polytechnical Institute, Lwow, 1953.
2. V. P. Sigorskii, Methods For Analyzing Electrical Circuits Containing MultipoleElements [in Russian] (Academy of Sciences Ukrainian SSR Press, 1958).
3. V. F. Vlasov, Electrovacuum Instruments [in Russian] (Communication, 1949).
4. A. M. Bonch-Bruevich, The Application of Electronic Tubes In Experimental Physics [in Russian] (State Power Engineering Press, 1955).
5. Ya. S. Itskhoki, Pulse Devices [in Russian], Soviet Radio Press (1959).
6. The Reliability Of Terrestial Radio-Electronic Equipment [in Russian] (Soviet Radio Press, 1957).
7. S. N. Krize, Low-Frequency Voltage Amplifiers [in Russian] (State Power Engineering Press, 1953).

PNEUMATIC INTEGRATOR WITH APERIODIC UNIT THROTTLE IN LOW PRESSURE AREA

Ch'i Ju-hsien (Moscow)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp. 1133-1138, August,1961
Original article submitted March 30, 1961

A pneumatic integrator with an aperiodic throttle in a low pressure area is described. Results of the experimental investigation of the integrator model are given. The most important factors influencing integrator accuracy are discussed.

Although we have an extensive, uniform system of nomenclature for pneumatic aggregates it does not include such an important device as the integrator. This omission is not accidental. It is due to the fact that when operating at working pressures of 0 - 1 atmospheric units the output characterisitics of the throttle are not linear. As a result of this we cannot obtain a purely aperiodic unit when using a dead chamber, and therefore we cannot build an accurate integrator. This is confirmed, for example, when we study the operation of the integrator built on the basis of the regulating block AUS 4RB-32A. This integrator is very inaccurate (the results contain large errors).

However at low pressures the characteristics are linear. This enabled Ferner [1], for example, to build an extremely accurate integrator which operated over the low pressure range (0 - 100 mm,water pressure). However, at present the overwhelming majority of industrial pneumatic devices operate at normal pressures. Therefore, only very limited application can be made of an integrator operating at low pressures.

Therefore,we carried on research in the Pneumohydraulic Automation Laboratory of the IAT AN SSSR devoted to the possibility of building an accurate integrator which would operate at normal pressures. We were able to do this by artificial operation of the aperiodic throttle at low pressures (0 - 200 mm,water pressure).

Analysis of the Integrator

The integrator circuit is shown in Fig. 1. Its basic elements consist of: the adder I, pressure reducer II (a pressure divider having a coefficient of amplification of 1/K), an amplifier III having a coefficient of amplification of unity, aperiodic unit IV,and the pressure amplifier V having a coefficient of amplification k.

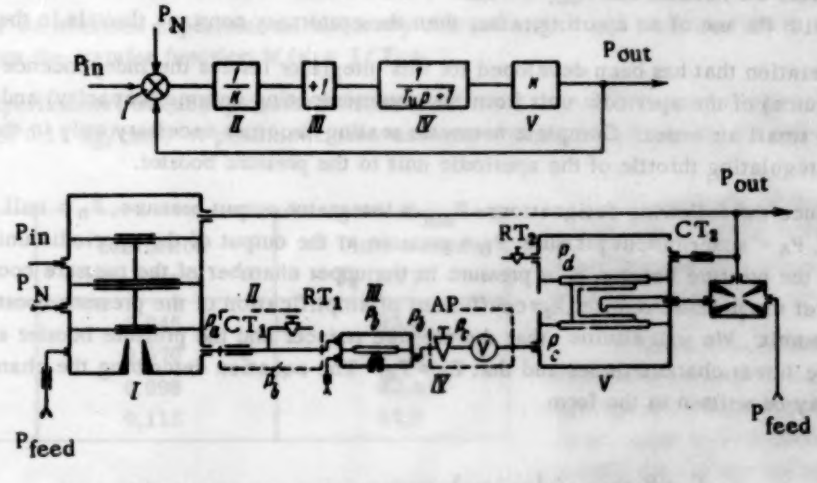


Fig. 1. Block diagram and schematic circuit of integrator.

The adder I consists of a standard block summing relay, a pneumatic aggregate system of the type BC-34A. There are three chambers in the block: two add with a plus sign and one with a minus sign. An input signal P_{in}

(Fig. 1) enters the lower plus chamber of the adder, a positive feedback signal P_{out} enters the upper plus chamber; P_{out} is obtained after passing through the amplifier; a pressure P_N which corresponds to the zero level is applied to the minus chamber.

The pressure reducer II represents a pneumatic flow-through chamber. At the output of the chamber there is a constant throttle CT_1 which limits the air input to the chamber; at the output there is a regulating throttle RT_1 which limits the air output to the atmosphere. Experimental studies show that for a choice of constant and regulating throttles we can obtain a practically linear dependence of the pressure inside the chamber P_b upon the pressure P_a in front of the constant throttle for k = 50 - 100. The constant throttle consists of a capillary 0.18 in diameter and 30 mm long, and the regulating throttle is in the form "cone-cone" with a conical angle of 4°. As a result of the small volume of the inter-throttle chamber the pressure reducer acts sufficiently rapidly.

Amplifier III, which has a transfer function $W_1(s) = 1$ is in the form of a follower or a "detecting" chamber. It operates under low pressure conditions; its membrane is made of a thin elastic. It serves as a separator or divider between the divider II and the aperiodic unit IV.

The aperiodic unit is a direct connection between the regulating throttle (pneumatic resistance) and the dead pneumatic chamber. A standard regulating throttle such as is found in the regulating AUS blocks for setting the isodrome time, is used to measure the time constant T_i of the aperiodic unit IV.

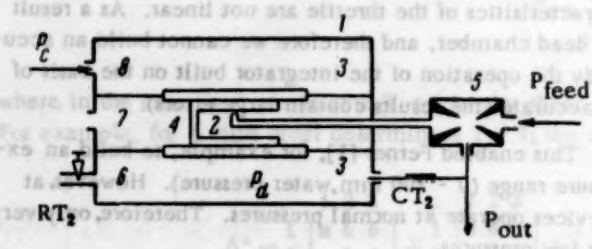


Fig. 2.

The last unit of the integrator V increases the pressure (it is an amplifier). Its construction, Fig. 2, is based upon the operational amplifier [2]. The unit is meant to raise the low pressure of the aperiodic unit (0 - 200 mm water) to normal pressure (0 - 1 atm.). This unit (pressure booster) consists of a pneumatic amplifier with a coefficient of amplification k_2 = 50 - 100. It consists of a body 1 containing a nozzle 2 and a membrane block with two identical thin membranes 3, whose stiff centers are connected by Π-shaped bracket 4 which acts as a slide (or shutter) for the nozzle. The pneumatic amplifier also contains a constant ejecting throttle 5. The membrane block forms three chambers 6, 7, and 8 inside the body 8. Chamber 8 is a dead chamber and chamber 6 is a pass-through chamber. The pass-through chamber 6 has a constant throttle CT_2 at its input and a regulating throttle RT_2 at its output. These throttles are identical to the corresponding throttles in the pressure reducer II (Fig. 1). When a pressure P_c is applied to chamber 8 the pressure in chamber 6 is P_d. The ejecting throttle 5 is applied to the input valve, with a pressure P_{feed} = 1.4 atm, and to the nozzle 2. The output pressure is built up in the inter-nozzle chamber of the ejector. It is also the output pressure of the integrator. Obviously $P_{out} = P_{feed}$ when the shutter (or slide) completely covers the nozzle, and P_{out} = 0 atm when the slide does not cover the nozzle at all. The latter condition is associated with the use of an ejecting, rather than the customary constant, throttle in the pressure booster.

The method of operation that has been developed for this integrator insures the independence of the working pneumatic capacity (volume) of the aperiodic unit from the communicating volume (capacity) and it also permits us to compensate for the small air losses. Complete hermetic sealing becomes necessary only in the part of the circuit extending from the regulating throttle of the aperiodic unit to the pressure booster.

We will now introduce the following designations: P_{out} = integrator output pressure, P_n = null pressure, P_{in} = integrator input pressure, P_a = adder output pressure, P_b = pressure at the output of the aperiodic unit, P_d = pressure in the lower chamber of the pressure booster, P_c = pressure in the upper chamber of the pressure booster, k_1 = coefficient of amplification of the pressure reducer, k_2 = coefficient of amplification of the pressure booster and T_i = time constant of the aperiodic unit. We will assume that the pressure reducer and the pressure booster are purely amplifying units, that they have linear characteristics, and that $P_c = P_d$. The equation describing the changes in the integrator output pressure may be written in the form

$$\frac{T_i}{k_2}\frac{dP_{out}}{dt} + \left(\frac{1}{k_2} - k_1\right) P_{out} = k_1 (P_{in} - P_N). \qquad (1)$$

For $1/k_2 = k_1 = k$, equation (1) becomes

$$T_i \frac{dP_{out}}{dt} = (P_{in} - P_N). \quad (2)$$

which is the equation of the integrator.

It has the solution

$$P_{out} = \frac{1}{T_u} \int_0^t (P_{in} - P_N)\, dt.$$

The transfer function of the integrator has the form

$$W(s) = \frac{1}{T_i s},$$

i.e., the integrator has the transfer function of an ideal integrating unit.

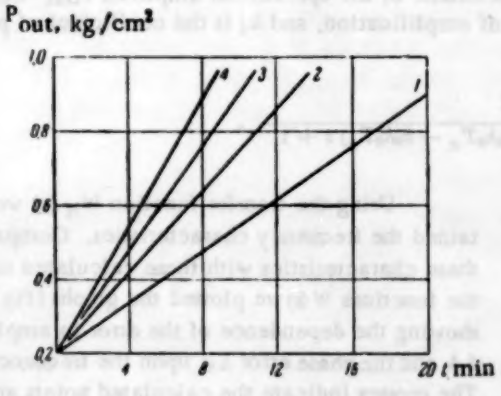


Fig. 3

Experimental investigations carried out using a model of the integrator enabled us to determine the linearity of the integrator characteristics and also the band of operating frequencies of the device.

First of all studies were made of the integrator linearity. The linearity criterion [3] was the equality of the time constants T_i for various types of unit steps (jumps). As examples, we present the curves of the integrator acceleration for various excitations 0.045, 0.07, 0.098, and 0.116 atm. The values of the excitation for various integrator time constants are given in the table.

We see that the values of the time constants do not differ by more than 0.5 - 1.0%; therefore, we can consider the integrator a linear integrator.

In order to determine the band of operating frequencies of the integrator we obtained experimental frequency characteristics and compared them with the characteristics computed from the transfer function $W(s) = 1/T_i s$.

The experimental frequency characteristics were taken over a frequency range of 0.07 - 6.8 rad/sec for an input signal of 0.12 kg/cm². A pneumatic sine-detector was used in obtaining the frequency characteristics.

No. of curve in Fig. 3	Excitation P_{in} (atm)	Time constant T_i (sec)
1	0.045	83.4
2	0.070	83.5
3	0.098	83.4
4	0.116	84.8

An example of the amplitude and phase characteristics is given in Fig. 4. The circles indicate experimental data while the crosses indicate the calculated values.

Comparison of the experimental and the calculated frequency characteristics shows that the experimental frequency characteristics give a sufficiently accurate approximation of the calculated characteristic for various values of T_i. However, the band of working frequencies decreases as the time constant T_i increases. We see from Fig. 4 that for the frequency region 0.07 - 5.5 rad/sec the experimental and calculated results coincide. With a further increase in frequency the experimental characteristics deviate from the calculated characteristics.

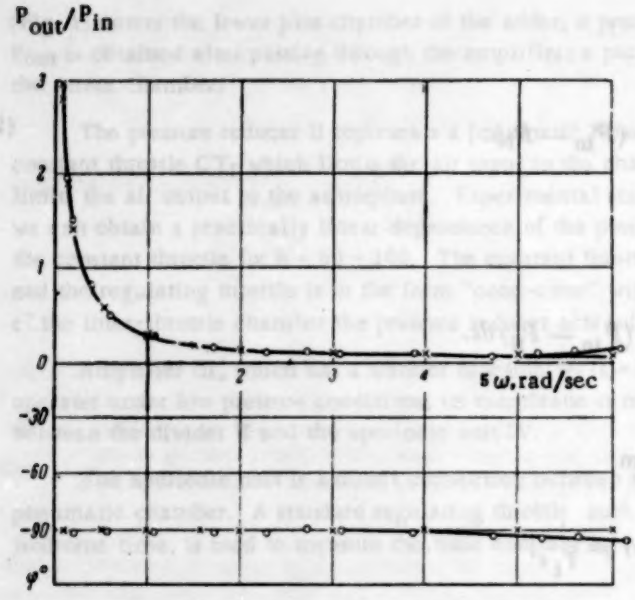


Fig. 4.

Comparison of the experimental frequency characteristics with the calculated characteristics shows that in the frequency range of 0 - 6.3 rad/sec the equation of motion of the integrator obeys the linear differential equation (2). When we increase the frequency further an error appears. An error analysis was made in an effort to try to discover the basic reason for the deviation of the dynamic characteristics of the real integrator from the characteristics of the ideal integrator.

We made the assumption that the greater part of the error appearing in the integrator operation was due to the amplifier (pressure booster) since actually P_c is not equal to P_d (the characteristics of the operational amplifier are not purely relay characteristics). Also the volume of chamber 6 (Fig. 2) affects the operation of the integrator. Because of the large volume of this chamber the pressure booster is not a purely amplifying unit. Therefore, we must take into account its time constant T_d.

The transfer function of the actual integrator which takes into account the booster in the aperiodic unit by means of the equation $T_d\ dP_d/dt + P_d = 1/k_3 P_{out}$ and the static characteristic of the operational amplifier $P_{out} = k_0(P_c - P_d)$ where T_d is the time constant of chamber 6, k_3 is the coefficient of amplification, and k_0 is the coefficient of proportionality, has the form

$$W_{\text{н}}(s) = k_3 k_0 \frac{T_d s + 1}{T_d T_{\text{н}} s^2 + (T_d + T_{\text{н}} + k_3 k_0 T_{\text{н}} - k_3 k_0 T_d)\, s + 1}.$$

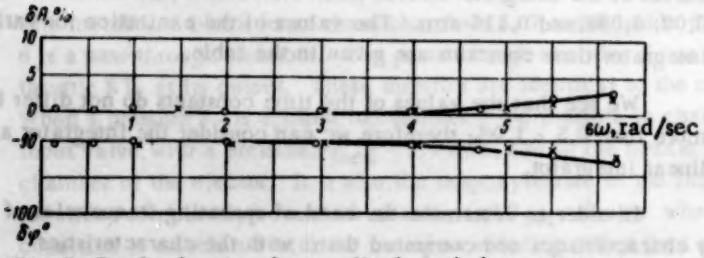


Fig. 5. Graphs showing the amplitude and phase errors as functions of frequency.

Using the transfer function $W_N(s)$ we obtained the frequency characteristics. Comparing these characteristics with those calculated using the functions $W(s)$ we plotted the graphs (Fig. 5) showing the dependence of the error in amplitude δA and the phase error $\delta\varphi$ upon the frequency. The crosses indicate the calculated points and the circles the experimental values. Therefore, we can consider that the presence of a finite volume in chamber 6 and the deviations of the operational amplifier from the relay characteristics result in the deviation of the dynamic characteristics of the actual integrator from the ideal integrator.

Therefore, we increase the frequency range of the integrator by decreasing the volume of chamber 6 and by improving the static characteristics of the amplifying unit.

CONCLUSION

Experimental tests of the pneumatic integrator with an aperiodic unit throttle in the low pressure area have led to the following conclusions.

1. The aperiodic unit throttle in the low-pressure area has enabled us to obtain integrator characteristics which possess good linearity.

2. The integrator allows a sufficiently large integration time.

3. The integrator operates over a sufficiently wide frequency range, and we can recommend its use in the automatization of industrial units.

LITERATURE CITED

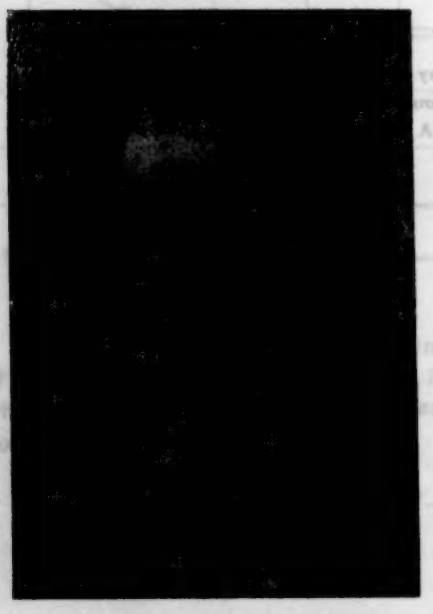
1. V. Femer, "Elements of pneumatic regulators," Coll.: Questions of Pneumo-Hydro Automation [in Russian], AN SSSR (1960).
2. G. T. Berezovets, V. N. Dmitriev, and A. A. Tal', A new system of pneumatic computing devices, Avtomatika i Telemekhanika 22, No. 1 (1961).
3. N. D. Lanin, "Small dimension continuous pneumatic computers having a time delay," Coll.: Questions of Pneumo-Hydro Automation [in Russian], AN SSSR (1960).

All abbreviations of periodicals in the above bibliography are letter-by-letter transliterations of the abbreviations as given in the original Russian journal. *Some or all of this periodical literature may well be available in English translation.* A complete list of the cover-to-cover English translations appears at the back of this issue.

A. A. ANDRONOV

(ON THE 60TH ANNIVERSARY OF HIS BIRTHDAY)

Translated from Avtomatika i Telemekhanika, Vol. 22, No. 8,
pp 1139-1142, August, 1961



Sometimes situations arise in the history of applied science and technology where the solution of the more realistic and pressing problems is retarded due to the seeming absence of adequate mathematical apparatus. However, it often turns out that the required mathematical methods have already been discovered but have been put away and forgotten. The uncovering of previously developed mathematical apparatus that enables scientists to solve important problems in physics and technology is a discovery of the first magnitude.

This, for example, was the situation which existed in the first half of the nineteenth century when B. Thompson (Lord Kelvin), having solved the problem of the propagation of electromagnetic oscillations, discovered the known but not previously used technique of Fourier series and thus laid the basis for the solution of many of the most important discoveries in electronics, radio, and oscillation theory.

A situation of this sort existed in science and technology at the end of the 1920's. With the invention of the electron tube, progress in radio and radiophysics was retarded by the absence of mathematical techniques that would have enabled scientists to describe and analyze the generation of oscillations in nonlinear systems. Primitive efforts to create a quasi-linear theory for vacuum tube generators satisfied neither the engineers nor the physicists and it seemed that there were no accurate methods for solving the nonlinear problems that arose because of the absence of existing methods for integrating nonlinear differential equations. On the other hand the well known mathematical work of Poincaré on the qualitative theory of differential equations and of Lyapunov on general stability theory could not be applied directly; they seemed to be digressive investigations which were of importance only in the general aspects of the development of mathematical theory.

A. A. Andronov was the first to understand the adequacy of Poincare's qualitative theory of differential equations and Lyapunov's work on the stability of motion of nonlinear oscillations. A. A. Andronov was then an aspirant at the University of Moscow. He was the first to indicate the adequacy of Poincaré's theory of limiting cycles to the solution of problems involving vacuum tube oscillators and the stability theory of Liapunov to problems involving the stability of generated oscillations. A. A. Andronov understood that the oscillations generated in nonlinear systems were a special class of oscillations and he proposed the term "self-oscillation" which was later used in science throughout the world.

A. A. Andronov's scientific efforts were then devoted to the study of self-oscillations and associated phenomena.

A. A. Andronov was the student and close co-worker of the famous Russian physicist, Academician L. I. Mandel'shtam and many of his students and co-workers. However, A. A. Andronov occupies a special and original place in the L. I. Mandel'shtam school. A large part of the work of the other scientists of this school was devoted to the discovery of new oscillatory phenomena by means of various parameter variations and the development of these methods, but A. A. Andronov devoted himself almost entirely to the study of the phenomena of self-oscillation by means of the qualitative methods of the theory of differential equations and the precise methods which are necessary in the development of the method of approximation (and summation).

A characteristic of A. A. Andronov's work was his ability in separating the many difficult problems which arose into particular abstract, narrow mathematical classes. These problems which were extremely important from the technical point of view could now be solved.

A typical example of the above type of problem were the class of dynamic systems which A. A. Andronov and L. S. Pontryagin called "coarse" systems. It turned out that practically all technical problems involved this type of problem and therefore by separating out this class of problem it was possible to cross off a large number of problems which were mathematically very difficult.

A. A. Andronov and A. A. Witte together developed the self-oscillatory theory of multivibrators. This laid the foundation for a series of works of the scientists of the Mandel'shtam school and for other investigations in the theory of discharge oscillations and a study of "discontinuity conditions".

Important discoveries were made by A. A. Andronov and A. A. Witte in the theory of the "capture" of the self-oscillation frequency by the perturbing force.

A. A. Andronov did not limit himself to applying the mathematical theory, which he had discovered but which had previously been developed by others, to physical oscillations. He, together with other mathematicians working under his supervision, made significant contributions to the further development of the qualitative theory, and in this the work of isolating the various classes of coarse systems played a decisive role.

In 1937 the theories developed by A. A. Andronov, his friends and co-workers were so far advanced with respect to systems having one degree of freedom that it was possible to publish the results in the form of a monograph. The book written by A. A. Andronov, A. A. Witte, and S. E. Khaikin, "Oscillation Theory",* soon became not only the outstanding reference book, used by everyone engaged in the study of oscillations, but also helped direct young people who were preparing to engage in scientific work in this field. Translated into the English language it soon received world wide recognition and led to the publication not only in Russia but all over the world of a number of works which led to further contributions to the theory of oscillations in the direction indicated by A. A. Andronov. At the same time it became clear to Andronov himself that he should not limit himself to systems with one degree of freedom. However, radio physics and radio engineering systems did not provide such convenient examples of systems with many degrees of freedom as they did of systems with one degree of freedom. A. A. Andronov came across Vishnegradov's classical problem which is associated with the theory of direct automatic regulation and which Zhukovski, Mises, Grdin, and many other leading scientists had previously tried to solve. A. A. Andronov understood that this problem might play the same roled in third order systems that the study of the vacuum tube generatory system played in second order systems. He devoted much time and effort to the solution of this classical problem in regulation theory. A. A. Andronov's and A. G. Maier's comprehensive solution of this problem serve, up to this time, as an example of the application of mathematics to the solution of concrete technical problems – it is precise, original, and bold in overcoming difficulties and still it is simple enough so that every engineer can understand the results.

After solving this problem a significant part of A. A. Andronov's work was devoted to the study of individual problems in the theory of automatic regulation which described or led to the investigation of low order systems. Even though each one of these problems was a realistic problem for that time, as far as the interests of science and of A. A. Andronov were concerned, these studies gradually led to the development of the mathematical apparatus requisite for a precise theory of nonlinear systems with several degrees of freedom. Here another discovery was made which was nearly equal in scientific value to that of the discovery made by Andronov when he was an aspirant (described above). Andronov and his co-workers found that the theory of pointwise transformations which up to that time had had no important technical applications was adequate for the solution of piecewise linear problems and permits us to convert the working engineering method – the method of successive approximations into a first class apparatus for scientific studies. In the hand of the A. A. Andronov school even during his lifetime the theory of pointwise transformation became, on the one hand, a tool that was used in investigation and with the aid of which a whole series of important scientific discoveries were made and, on the otherhand, a tool used in engineering investigation in those cases where the approximate methods were not adequate and it was necessary to obtain an exact solution of the problem.

* Through no fault of A. A. Andronov and S. E. Ehaikin, A. A. Witte was not listed in 1937 as one of the co-authors of the monograph, and it is known throughout the world as: "Oscillation Theory" by A. A. Andronov and S. E. Khaikin. However, when the second edition was published in 1960 A. A. Witte's name was added to the list of authors.

The investigations of dynamic systems made by A. A. Andronov and his students of the dynamics of various different types of physical systems from a common viewpoint and also of the separation of the various classes of coarse systems naturally led A. A. Andronov to try to formulate a general theory of machine dynamics and in particular to obtain a scientific formulation for the various machines.

Time has shown that A. A. Andronov's unexpected death did not stop further work in the direction which he had indicated. His students and co-workers who are grouped mainly in the city of Gorky – they are often called the Gorky school of oscillation and radiophysics - and also many coworkers and students of A. A. Andronov working in other cities and institutes are successfully continuing his investigations.

A. A. Andronov was a deputy to the Upper Soviet of the USSR, a member of the Presidium of the Upper Soviet of the RSFSR. His relation to his duties as a deputy was exemplary.

The students, friends, and all others who were fortunate enough to have known A. A. Andronov remember him not only as a remarkably fine scientist but also as a person of crystalline honesty and of high principles, as an active patriot as well as an outstanding worker and scientist.